Molecular phylogeny of frogs (Amphibia: Anura) based on complete mitochondrial genomes and partial nuclear genes

Iker Irisarri Aedo

UNIVERSIDAD AUTONOMA DE MADRID

Universidad Autónoma de Madrid
Facultad de Ciencias – Departamento de Biología

Molecular phylogeny of frogs (Amphibia: Anura) based on complete mitochondrial genomes and partial nuclear genes

Iker Irisarri Aedo
Madrid, 2012

Departamento de Biodiversidad y Biología Evolutiva
Museo Nacional de Ciencias Naturales
Consejo Superior de Investigaciones Científicas

Molecular phylogeny of frogs (Amphibia: Anura) based on complete mitochondrial genomes and partial nuclear genes

Memoria presentada por IKER IRISARRI AEDO para optar al grado de Doctor en Ciencias Biológicas

VºBº del director de tesis
Dr. Rafael Zardoya San Sebastián

VºBº del tutor de tesis
Dr. José Luis Bella Sombría

Madrid, 2012

Izarren hautsa egun batean bilakatu zen bizigai,
hauts hartatikan uste gabean noizpait ginaden gu ernai.
Eta horrela bizitzen gera sortuz ta sortuz gure aukera
atsedenik hartu gabe: lana eginaz goaz aurrera
kate horretan denok batera gogorki loturik gaude.

Gizonen lana jakintza dugu: ezagutuz aldatzea,
naturarekin bat izan eta harremanetan sartzea.
Eta indarrak ongi errotuz, gure sustraiak lurrari lotuz,
bertatikan irautea: ezaren gudaz baietza sortuz,
ukazioa legetzat hartuz beti aurrera joatea.

Gu sortu ginen enbor beretik sortuko dira besteak,
burruka hortan iraungo duten zuhaitz-ardaxka gazteak.
Beren aukeren jabe eraikiz ta erortzean berriro jaikiz
ibiltzen joanen direnak: gertakizunen indar ta argiz
gure ametsa arrazoi garbiz egiztatuko dutenak.

X. Lete, M. Laboa.

Y tú, juventud estudiosa, esperanza de nuestra renovación, que te consagras al trabajo en estos luctuosos días de nuestra decadencia, no te desalientes.

S. Ramón y Cajal.

Un día el polvo estelar se convirtió en germen de vida,
y de aquel polvo inesperadamente surgimos nosotros.
Y así vivimos, creando nuestras oportunidades
sin tregua: avanzamos por medio del trabajo
todos a una profundamente unidos en esa cadena.

Del trabajo humano, la sabiduría: transformar conociendo
ser uno con la naturaleza y establecer vínculos con ella.
Y afianzando las fuerzas, uniendo nuestras raíces a la tierra,
permanecer en ella: creando la afirmación de la negación,
tomando la negación como ley, adelante siempre.

Del mismo tronco del que surgimos, surgirán los otros,
brotes jóvenes que perpetuarán esta lucha.
Erigiéndose dueños de su encrucijada, levantándose al caer
irán caminando: por la fuerza y evidencia de los hechos
convertirán, con honestidad, nuestro sueño en realidad.

X. Lete, M. Laboa.

Eta zu, gaztedi ikastun , gure berritzearen itxaropena,
gure gainbeheraren egun hauetan lanari eskaintzen zarena,
ez zaitez etsitu.

S. Ramón y Cajal.

The dust of stars became one day the seeds of life,
from that dust unexpectedly we were born.
And so we live, creating our opportunities
with no rest: we advance by means of labour
being tightly bound together in this chain.

From human labour, knowledge: transforming by experience
to be one with nature, create links with it.
Strengthening our forces, linking our roots to the Earth,
long-lasting on it: creating the assertion of denial,
taking denial as law, always moving forward.

From the same trunk we emerged, others will emerge,
new shoots that will perpetuate this fight.
Gaining awareness of opportunities, getting up after falling
continuing to walk by the strength and evidence of the facts,
with trustworthiness converting our dream into reality.

X. Lete, M. Laboa

And you, studious youth, hope of our renovation, devoted
to work in these gloomy days of our decline, do not become
discouraged.

S. Ramón y Cajal.

Acknowledgements
Agradecimientos

Mi primer y más sincero agradecimiento es para Rafael Zardoya, director de esta tesis doctoral. Gracias por la confianza que has depositado en mí desde el inicio, permitiéndome descubrir más y más detalles de este mundo fascinante que es la evolución. Ha sido un verdadero privilegio compartir estos años, que me han instruido (creo) en la paciencia y tenacidad necesarias para alcanzar cualquier objetivo, así como en la necesidad de ser meticuloso (y pragmático) en el trabajo. Debo darte las gracias también por la libertad de la que he gozado para aprender y para equivocarme. Gracias por estar pendiente de mis preguntas y disponible a pesar de tu agenda; gracias por esos cinco minutos en forma de charla o email que tantas veces me han aclarado y orientado.

Deseo agradecer a mi tutor José Luis Bella su continuo apoyo y estímulo en este proyecto. Gracias por tus consejos y por facilitarme el trabajo académico y burocrático.

Quiero dar las gracias en especial a aquéllos directamente involucrados en esta tesis y cuyo trabajo se refleja en ella. La participación y el trabajo de Miguel Vences ha sido crucial en muchos aspectos teóricos y prácticos; quizá el más patente sean las tinciones anatómicas de las laringes. Gracias Miguel por transmitirme tu pasión por la ciencia, la evolución y los anfibios. Gracias además por haberme acogido en la Technische Universität Braunschweig; ha sido un verdadero placer y privilegio trabajar contigo y tu grupo. A Diego San Mauro, muchas gracias por guiarme en mis primeros pasos en el laboratorio, instruirme en los análisis filogenéticos y aleccionarme en tantas y tantas cuestiones. Muchísimas gracias a Federico Abascal, por colaborar en los análisis, pero sobre todo por enseñarme tantas y tantas cosas, confiar en mí y animarme. Gracias por tu amistad. Many thanks to Frank Glaw (Zoologische Staatssammlung München) for the original suggestion of studying sound production in *Pseudhymenochirus*, as well as for recording it. Thanks also for your insightful comments and support, which have been crucial for the successful end of this thesis. I am grateful to David M. Green (Redpath museum, McGill University) and Annemarie Ohler (Muséum national d'Histoire naturelle) for providing critical samples and bright comments.

I would also like to acknowledge the people who provided tissue samples or assisted in fieldwork: A. Crottini, R. Boistel, M. Burger, A. Channing, S. Esser, M. García-París, K. Kunz, B. Love, Í. Martínez-Solano, G. Parra-Olea, and S. Richards. Particularly, I would like to thank Angelica Crottini for her friendship and all the good times spent together, as well as for sharing her knowledge and enthusiasm.

I am extremely grateful to Lars S. Jermiin, who kindly accepted me in CSIRO Ecosystem Sciences at Canberra, and introduced me to the fascinating world of evolutionary model assumptions. I have been lucky to learn many good qualities from you, including excellent scientific practice. My gratitude also goes to Douglas J. Eernisse; many thanks for the warm welcome to the California State University at Fullerton, and for discovering me lots of wonderful things, including the amazing world of chitons, and many incredible spots in California. Doug, you rock!

Muchísimas gracias a Mario García-París, David Buckley, Íñigo Martínez-Solano, Ernesto Recuero y otros habitantes que han pasado por el Z-418. Habéis sido un apoyo muy importante durante el desarrollo de esta tesis. Mil gracias por vuestros consejos, y por estar siempre dispuestos a escucharme y discutir conmigo sobre ranas, evolución, ciencia, o cualquier otra cosa. Sois una fuente inagotable de sabiduría, y además, muy buena gente, ¡gracias!

Quiero mostrar mi agradecimiento especialmente a mis amigos y compañeros de fatigas del Museo. A David Osca y Jorge Gutiérrez, por estar siempre dispuestos a ayudar y soportar mi falta de paciencia. A Raquel Álvarez, por traer la alegría; te mereces lo mejor. A Silvia Perea y Lourdes Alcaraz, por escucharme, apoyarme, y enseñarme tantas y tantas cosas. Muchas gracias, Silvia, por todo tu apoyo durante las fases finales de esta tesis. A Fernando Alda, Patricia Cabezas, Elena G. González, Patricia Ornelas y Carlos Pedraza, por vuestros sabios consejos y por estar siempre dispuestos a echar una mano con la mejor de vuestras sonrisas. A Anna M. Addamo y Carlos Toledo, por el año compartido. A Melinda Hofmann, por revisar mi traducción al inglés de Izarren hautsa. Al resto de becarios y gentecilla que han pasado por el Museo: Pedro Abellán, Iván Acevedo, Pau Aleixandre, Anabela Arraiol, Miriam Casal, Diana Delicado, Saioa Fernández-Beaskoetxea, Fernando A. Fernández-Álvarez, Teresa García, Virginia González-Jimena, Amparo Hidalgo, Ainhoa Iraola, Diego Llusia, Neus Marí-Mena, Paloma Mas, Bárbara Martín-Beyer, Pilar Pavón, Marisa Peláez, Mari J. Ruiz-López, Gregorio Sánchez, Luis M. San-José, Chiara Settani, Juanes Uribe, Alberto Vicens (Chechu), etc. Seguro que me dejo a unos cuantos... Gracias de todo corazón por vuestra amistad, los buenos momentos y las fiestecillas.

Así mismo, me gustaría agradecer a Marta Barluenga, Alexandra Cieslak, Regina Cunha, Ignacio Doadrio, Patrick S. Fitze, Cristina Grande, Noemí Guil, Annie Machordom, Isabel S. Magalhaes, Borja Milá, Carolina Noreña, Borja Sanchíz y José Templado, el haber compartido conmigo sus conocimientos, consejos y anécdotas. Gracias al resto de los compañeros del laboratorio (incluyendo a Piluchi, Ricardo e Isabel) por vuestra ayuda y por amenizar el quehacer diario. Me gustaría recordar aquí, especialmente, a Marina Alcobendas, una magnífica compañera a la que echamos de menos.

Por último, agradezco a mi familia que me apoya siempre, en todo. A mis padres José Luis y Sabi. A mi hermano Oier, que me ha echado un cable con el diseño de esta tesis. A Jose, gracias por estar siempre ahí. A mis abuelos Julián, María Jesús, Agustín y Rosario, por vuestro ejemplo.

No quisiera olvidarme de otros amigos que me han acompañado (¡en especial los de Iruñea!). Menos mal que me recordáis que hay vida más allá de las puertas del Museo. Gracias por todos los buenos ratos.

Esta tesis ha sido posible gracias a una beca predoctoral JAE del Consejo Superior de Investigaciones Científicas (CSIC) cofinanciada por el Fondo Social Europeo (ESF) y la financiación por parte del Ministerio de Ciencia e Innovación, a través de proyectos de investigación a Rafael Zardoya (CGL2007-60954 y CGL2010-18216).

Table of contents

List of abbreviations

Nucleotides and amino acids are referred according to the IUPAC, chemical elements use conventional symbols, and units of measurement follow the International System. Names of genes are in lowercase italics, and names of gene products are in regular uppercase.

Next are other non-conventional abbreviation used throughout the text (names of mitochondrial genes follow Boore, 1999):

AIC	Akaike information criterion
AU	approximately unbiased
atp6	subunit 6 of the ATP synthase
atp8	subunit 6 of the ATP synthase
bdnf	brain-derived neutrophic factor
BIC	Bayesian information criterion
bp	base pair
CI	credibility interval
cob	cytochrome b
cox1	unit 1 of the cytochrome oxidase
cox2	unit 2 of the cytochrome oxidase
cox3	unit 3 of the cytochrome oxidase
CR 5' region	region of the mt genome including the genes *nad6, trnE, cob, trnT,* and *trnP* plus the control region
CSB	conserved sequence block
cxcl12	ligand 12 of the chemokine family (= stromal-derived factor 1; SDF-1)
cxcr4	chemokine receptor type 4
ΔAIC	difference between AIC values
Γ	gamma distribution
GTR	general time-reversible model of nucleotide substitution
I	proportion of invariable sites
IUCN	International Union for Conservation of Nature
LTPF	tRNA gene cluster containing the *trnL–(CUN), trnT, trnP,* and *trnF* genes
-*ln*L	natural logarithm of the likelihood
LRT	likelihood ratio test
Mb	megabase pair (106 bp)
MCMC	Markov chain Monte Carlo
mt	mitochondrial
mtREV	reversible model of amino acid replacement for mitochondrial proteins
mya	million years ago

nad1	NADH dehydrogenase subunit 1
nad2	NADH dehydrogenase subunit 2
nad3	NADH dehydrogenase subunit 3
nad4	NADH dehydrogenase subunit 4
nad4L	NADH dehydrogenase subunit 4L
nad5	NADH dehydrogenase subunit 5
nad6	NADH dehydrogenase subunit 6
PCR	polymerase chain reaction
pomc	proopiomelanocortin
rag1	recombination-activating gene 1
rag2	recombination-activating gene 2
rho	rhodopsin
rRNA	ribosomal ribonucleic acid
rrnL	large subunit of the rRNA gene
rrnS	small subunit of the rRNA gene
SD	standard deviation
SDF-1	stromal cell-derived factor 1 (= *cxcl12*)
slc8a1	member 1 of the solute carrier family 8 (also known as *ncx1*)
slc8a3	member 3 of the solute carrier family 8 (also known as *ncx3*)
TAS	termination-associated sequences
tRNA	transfer ribonucleic gene
trnA	transfer RNA gene for alanine
trnC	transfer RNA gene for cysteine
trnD	transfer RNA gene for aspartic acid
trnE	transfer RNA gene for glutamic acid
trnF	transfer RNA gene for phenylalanine
trnG	transfer RNA gene for glycine
trnH	transfer RNA gene for histidine
trnI	transfer RNA gene for isoleucine
trnK	transfer RNA gene for lysine
trnL– (CUN)	transfer RNA gene for leucine that recognizes the codon family CUN
trnL– (UUR)	transfer RNA gene for leucine that recognizes the codon family UUR
trnM	transfer RNA gene for methionine
trnN	transfer RNA gene for asparagine
trnP	transfer RNA gene for proline
trnQ	transfer RNA gene for glutamine
trnR	transfer RNA gene for arginine
trnS–(AGY)	transfer RNA gene for serine that recognizes the codon family AGY
trnS–(UCN)	transfer RNA gene for serine that recognizes the codon family UCN
trnT	transfer RNA gene for threonine
trnV	transfer RNA gene for valine
trnW	transfer RNA gene for tryptophan
trnY	transfer RNA gene for tyrosine
WANCY	tRNA cluster including *trnW, trnA, trnN, trnC,* and *trnY*, as well as the origin of replication of the light strand

Summary

Modern anurans (frog and toads) have a long evolutionary history of more than 200 million years, having undergone an extraordinary phylogenetic and ecological diversification, which gave rise to the 6,000 species currently recognized. They exhibit a variety of morphological, ecological and behavioural adaptations that allowed them to inhabit many different environments, across every continent (except Antarctica) and most continental islands. Early studies on morphology made important contributions towards understanding the evolutionary history of frogs. Later on, the analysis of sequence data, and the development of powerful and sophisticated probabilistic methods of inference provided substantial advance in the phylogeny of frogs. In the last decade, the application of molecular phylogenetic techniques allowed resolving some of the controversial issues regarding higher-level relationships among living frogs, and currently, we have a reasonably well-defined portrait of the frog tree of life. Yet, many important issues still remain under debate. Most studies agree on a monophyletic origin of the Anura, and four major lineages are generally recognized (Discoglossoidea, Pipoidea, Pelobatoidea, and Neobatrachia), along with two basal genera (*Leiopelma* and *Ascaphus*) of uncertain placement.

Here, a sequence data set combining complete mitochondrial genomes and nine nuclear loci was used to estimate a molecular phylogeny of frogs and tackle long-standing contentious questions. First, the relative phylogenetic position of *Leiopelma* and *Ascaphus*, two genera that have traditionally been considered basal within Anura, was addressed. Second, new sequence data was generated for Pipoidea, aiming to discriminate among competing hypotheses for the relationship between this group and Discoglossoidea, as well as to elucidate the internal phylogeny of the family Pipidae. Taking advantage of having reconstructed a robust phylogenetic framework for Pipoidea, the unusual sound production mechanism of the poorly known species *Pseudhymenochirus merlini* was studied by means of behavioural and anatomical observations. Third, the phylogenetic relationships among basal families within Neobatrachia were examined to further understand their extraordinary diversification, as well as the evolution of the specific gene order and higher substitution rates of neobatrachian mitochondrial genomes.

The sequence data set combining complete mitochondrial genomes and nine nuclear loci showed good phylogenetic performance in inferring deep level phylogenetic relationships among frogs. Results from phylogenetic analyses pointed to a sister group relationship between *Leiopelma* and *Ascaphus* (Amphicoela hypothesis), and their placement together as sister group of all remaining frogs. Remarkably, the mitochondrial genome of *Leiopelma archeyi* has a gene

arrangement that is unique among frogs but convergent with that of other vertebrates. Moreover, comparative data on mitochondrial gene orders from other vertebrates strongly suggested that the 5' end of the control region is a hot spot of gene rearrangement.

Further phylogenetic analyses were congruent in supporting the successive branching of the five major lineages of living frogs, as (i) Amphicoela, (ii) Discoglossoidea, (iii) Pipoidea, (iv) Pelobatoidea, and (v) Neobatrachia. Within Pipidae, both mitochondrial and nuclear data were congruent in recovering a deep divergence between an American lineage (*Pipa*) and an African lineage, in which dactylethrines (*Xenopus* + *Silurana*) were the sister group of hymenochirines (*Hymenochirus* + *Pseudhymenochirus*). Behavioural observations on *Pseudhymenochirus* unambiguously showed an air-driven mechanism for sound production, in contrast to all other members of the family, which have a mechanism independent of air. Given the derived phylogenetic position of *Pseudhymenochirus*, this observation was interpreted as a reversal to the ancestral non-pipid condition, which according to the performed anatomical observations, seems to have evolved constrained under the restrictions imposed by the derived larynx of pipids.

New mitogenomic and nuclear data on basal neobatrachian families provided insights into their phylogenetic position. *Heleophryne* was recovered as the sister group to all other neobatrachians. *Lechriodus* and *Calyptocephalella* were recovered as sister taxa, and both as sister to Nobleobatrachia. Within Nobleobatrachia, *Duttaphrynus* and *Telmatobius* were sister genera to the exclusion of *Hyla*. Phylogenetic analyses also suggested a sister group relationship between *Sooglossus* and Ranoides. The analysis of mitochondrial genomes within a phylogenetic framework provided further information to understand gene rearrangement dynamics and unravel mechanisms of molecular evolution. The reconstructed phylogeny showed that the neobatrachian-specific mitochondrial gene order was already present in the earliest branching living lineage of the group, suggesting that it might represent a molecular synapomorphy for Neobatrachia. Furthermore, mitochondrial substitution rates were found to be accelerated at the origin of Neobatrachia, and were higher in both basal and derived neobatrachian lineages compared to non-neobatrachian frogs. However, no consistent patterns were found among the nine nuclear genes studied. Further examination of mitochondrial protein-coding genes suggests that relaxation of purifying selection might account, at least in part, for the observed rate acceleration at the origin of the Neobatrachia.

Resumen

Los anuros actuales (ranas y sapos) tienen una larga historia evolutiva de más de 200 millones de años, durante la que han sufrido una extraordinaria diversificación filogenética y ecológica que ha dado lugar a las 6.000 especies reconocidas actualmente. Presentan un abanico de adaptaciones morfológicas, ecológicas y de comportamiento que les han permitido habitar ambientes muy dispares en todos los continentes (excepto en la Antártida) y en la mayoría de islas continentales. Los estudios tempranos sobre morfología contribuyeron notablemente a la comprensión de la historia evolutiva de los anuros. Más tarde, el análisis de datos de secuencia y el desarrollo de métodos probabilísticos de inferencia más poderosos y sofisticados provocaron un avance sustancial en el conocimiento de su filogenia. Durante la última década, la aplicación de las técnicas de filogenética molecular ha permitido resolver algunas de las cuestiones controvertidas respecto a las relaciones entre los grandes linajes de anuros actuales y, a día de hoy, tenemos una imagen razonablemente bien definida del árbol de la vida de los anuros. Sin embargo, varias cuestiones importantes siguen aún siendo motivo de debate. La mayoría de los estudios están de acuerdo en el origen monofilético de Anura, dentro del cual se reconocen generalmente cuatro grandes linajes (Discoglossoidea, Pipoidea, Pelobatoidea, y Neobatrachia), además de dos géneros basales (*Leiopelma* y *Ascaphus*) de posición incierta.

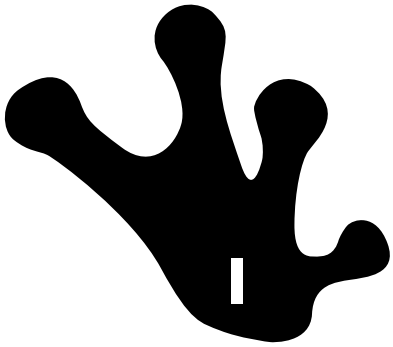
En la presente tesis se generó un conjunto de datos que combina genomas mitocondriales completos con nueve loci nucleares, y se utilizó para estimar una filogenia molecular de los anuros y abordar cuestiones conflictivas que permanecen aún sin resolver. En primer lugar se ha tratado de resolver la posición filogenética relativa de *Leiopelma* y *Ascaphus*, dos géneros tradicionalmente considerados basales dentro de Anura. En segundo lugar se ha generado nueva información molecular para Pipoidea, con el objetivo de discriminar entre las hipótesis de la relación entre este grupo y Discoglossoidea, propuestas anteriormente en la literatura, así como para elucidar la filogenia interna de la familia Pipidae. Tomando como base el marco filogenético robusto generado, se estudió el inusual mecanismo de producción sonora de la poco conocida especie *Pseudhymenochirus merlini*, por medio de observaciones de comportamiento y anatómicas. En tercer lugar se examinaron las relaciones entre las familias basales dentro de Neobatrachia, para poder así entender la extraordinaria diversificación de este grupo, así como para comprender la evolución del particular orden génico y las elevadas tasas sustitutivas características de los genomas mitocondriales de neobatráceos.

El conjunto de datos que combina genomas mitocondriales completos y nueve loci nucleares presenta un buen rendimiento para inferir relaciones filogenéticas profundas entre los anuros.

Los resultados de los análisis filogenéticos apuntaron a una relación de taxones hermanos entre *Leiopelma* y *Ascaphus* (hipótesis Amphicoela), y la posición de ambos como el grupo hermano de todo el resto de anuros. Curiosamente, el genoma mitocondrial de *Leiopelma archeyi* tiene un reordenamiento génico único entre los anuros, pero que al mismo tiempo es convergente con el de otros vertebrados. Es más, datos comparativos del orden génico mitocondrial de otros vertebrados sugieren claramente que la región 5' final de la región control es un punto caliente para la reorganización génica.

Los análisis filogenéticos posteriores fueron congruentes en apoyar la ramificación sucesiva de los cinco grandes linajes de anuros actuales, como (i) Amphicoela, (ii) Discoglossoidea, (iii) Pipoidea, (iv) Pelobatoidea, y (v) Neobatrachia. Dentro de Pipidae, tanto los datos mitocondriales como nucleares fueron congruentes en apoyar una divergencia temprana entre un linaje americano (*Pipa*) y otro africano, donde los dactylethrines (*Xenopus* + *Silurana*) fueron el grupo hermano de los hymenochirines (*Hymenochirus* + *Pseudhymenochirus*). Las observaciones del comportamiento en *Pseudhymenochirus* mostraron de modo inequívoco un mecanismo de producción sonora basado en aire. Dada la posición filogenética derivada de *Pseudhymenochirus*, esta observación fue interpretada como una reversión al estado ancestral no-pípido, pero que de acuerdo con las observaciones anatómicas realizadas, parece haber evolucionado constreñido bajo las restricciones impuestas por la laringe derivada de los pípidos.

Los nuevos datos mitogenómicos y nucleares para las familias basales dentro de Neobatrachia permitieron comprender mejor sus respectivas posiciones filogenéticas. *Heleophryne* fue el grupo hermano de todo el resto de neobatráceos. *Lechriodus* y *Calyptocephalella* fueron apoyados como taxones hermanos, y ambos como el grupo hermano de Nobleobatrachia. Dentro de Nobleobatrachia, *Duttaphrynus* y *Telmatobius* fueron géneros hermanos, excluyendo al género *Hyla*. Los análisis filogenéticos también sugirieron una relación de grupos hermanos entre *Sooglossus* y Ranoides. Basándose en el marco filogenético anterior se estudiaron los genomas mitocondriales para extraer nueva información que permitiera entender la dinámica de reordenamiento génico y desentrañar los mecanismos de evolución molecular. La filogenia obtenida mostró que el orden génico mitocondrial específico de neobatráceos estaba ya presente en el primer linaje actual del grupo, lo que podría sugerir que este ordenamiento representa una sinapomorfía molecular para Neobatrachia. Además, se demostró que las tasas sustitutivas mitocondriales se aceleraron en el origen de Neobatrachia, permaneciendo elevadas tanto en los linajes de neobatráceos más basales como en los más derivados, con respecto al resto de anuros no-neobatráceos. Sin embargo, no se encontró un patrón consistente entre los nueves genes nucleares estudiados. Tras examinar en profundidad los genes mitocondriales codificantes para proteínas, se encontraron evidencias de la relajación de la selección purificadora, lo cual podría explicar, al menos parcialmente, la aceleración de tasas observada en el origen de Neobatrachia.



INTRODUCTION

1.1. Anura: origin and evolution

Amphibians possess a long evolutionary history of more than 360 million years that began in the Late Devonian with the evolution of forms such as *Ichthyostega* and *Acanthosega* adapted to life in land from sarcopterygian fishes such as *Panderichthys* (Carroll, 2009). By the Late Carboniferous, early tetrapods had already extensively radiated into many different lineages including that of the ancestors of modern amphibians and amniotes, but most of these became extinct in the end-Permian mass extinction (Carroll, 2009). Living amphibians or Lissamphibia include three groups: Gymnophiona (caecilians), Caudata (salamanders and newts), and Anura (frogs and toads) (Fig. 1.1) (Duellman and Trueb, 1986). Tracing the origin of modern amphibians from Palaeozoic forms has proven difficult because fossil evidence is scarce and living amphibians possess markedly different morphology and, probably lifestyles, compared to any Palaeozoic lineage (Carroll, 2009). Therefore, three competing hypotheses have been proposed for the origin of extant amphibians: (i) that they are a monophyletic group that arose from Temnospondyli (Milner, 1988; Benton, 1990; Ruta et al., 2003; Ruta and Coates, 2007; Anderson et al., 2008; Sigurdsen and Green, 2011), (ii) that they are a monophyletic group that arose from Lepospondyli (Laurin and Reisz, 1997; Laurin, 1998; Vallin and Laurin, 2004; Pyron, 2011), and (iii) that they have a diphyletic origin with Anura and Caudata arising form temnospondyl dissorophoids, and Gymnophiona from lepospondyl microsaurs (Carroll, 2001; Schoch and Carroll, 2003; Carroll et al., 2004; Carroll, 2009). A single origin of living amphibians from temnospondyl ancestor is most widely accepted (Sigurdsen and Green, 2011), though the alternative single origin from Lepospondyli has not been unequivocally rejected (Carroll, 2009; Pyron, 2011). The possibility that Lissamphibia is not monophyletic is only marginally supported.

The only link between Palaeozoic and Mesozoic amphibians is *Gerobatrachus*, an Early Permian (*ca.* 280 million years ago [mya]) temnospondyl that seems to be closest to salamanders and frogs than to caecilians (Anderson et al., 2008). However, some authors (Marjanović and Laurin, 2008, 2009) have recently questioned its affinities, suggesting that it may be a temnospondyl with some lissamphibian convergences. In such a case, the oldest fossils unambiguously assigned to modern amphibians date back to the Early Triassic, and are represented by *Triadobatrachus* (Rage and Roček, 1989) and *Czatkobatrachus* (Evans and Borsuk-Bialynicka, 1998). Both show clear anuran affinities (Fig. 1.2) (Roček, 2000), even though they are not considered part of the crown-group, but rather stem-anurans (included in Salientia) (Carroll, 2009; Martín and Sanchíz, 2010). The oldest fossil for crown-group Anura is *Prosalirus*, which dates back to the Early Jurassic (Shubin and Jenkins, 1995). It retains some features from primitive tetrapods, but at the same time,

it shares many derived characters with basal living frogs, such as a modern configuration of the lower jaw, or notochordal and amphicoelous vertebrae (Shubin and Jenkins, 1995). *Vieraella* is slightly younger and possesses 10 notocordal amphicoelous presacral vertebrae (Fig. 1.2) (Roček, 2000). The oldest fossils ascribed to crown-group Caudata are *Iridotriton* (incertae sedis) (Evans et al., 2005b) and *Chunerpeton* (Cryptobranchidae) (Gao and Shubin, 2003) from the Later Jurassic, with an existing debate on which is oldest (Wang and Evans, 2006; Roelants et al., 2007). The earliest caecilian fossil is *Eocaecilia* from the Early Jurassic (Jenkins and Walsh, 1993), although it is probably not a member of the crown-group but rather a stem-caecilian (Carroll, 2000). The overall configuration of the skull and the lower jaw is typical of caecilians, but it retains a primitive configuration of the vertebrae and hind limbs, which are absent in all living members of the group (Carroll, 2009).

It is remarkable that a long gap in the fossil record of about 87 millions of years separates mid-Permian ancestors from modern caecilian, salamander and anuran crown-groups (Carroll, 2009). The absence of fossils that could represent intermediate stages during this long period has hampered the establishment of phylogenetic relationships among the three main lineages of living amphibians with certainty, given that first fossils known in each lineage possess basically the same body plan as their living descendents, and thus, they provide little more evidence as to their ancestry than do contemporary species (Schoch and Milner, 2004; Carroll, 2009). *Gerobatrachus* and Triassic salientians are the only exception, but specially *Triadobatrachus* and *Vieraella* are morphologically very close to crown-group frogs (Fig. 1.2) (Carroll, 2009). Therefore, it is likely that all three lineages of living amphibians acquired their specialized morphology very early in their evolutionary history (Zardoya and Meyer, 2001).

Despite the general agreement of a common origin for caecilians, salamanders and anurans, the long morphological gap separating both living and available fossil forms hindered the phylogenetic relationships among all three groups, and thus two competing hypotheses have been suggested. The sister group relationships between frogs and salamanders to the exclusion of caecilians ("Batrachia" hypothesis; Milner, 1988) is most widely accepted, and supported by both morphological evidence (Rage and Janvier, 1982; Duellman and Trueb, 1986; Milner, 1988; Trueb and Cloutier, 1991), as well as the most recent molecular studies (Zardoya and Meyer, 2001; San Mauro et al., 2005; Zhang et al., 2005a; Roelants et al., 2007; San Mauro, 2010). However, some initial molecular studies based on (mainly partial) sequences of mt or nuclear ribosomal RNA (rRNA) genes (Hedges et al., 1990; Hedges and Maxson, 1993; Hay et al., 1995; Feller and Hedges, 1998) recovered a sister group relationship between salamanders and caecilians, to the exclusion of frogs ("Procera" hypothesis; Feller and Hedges, 1998). The possibility of a sister group relationship between caecilians and frogs has never been proposed.

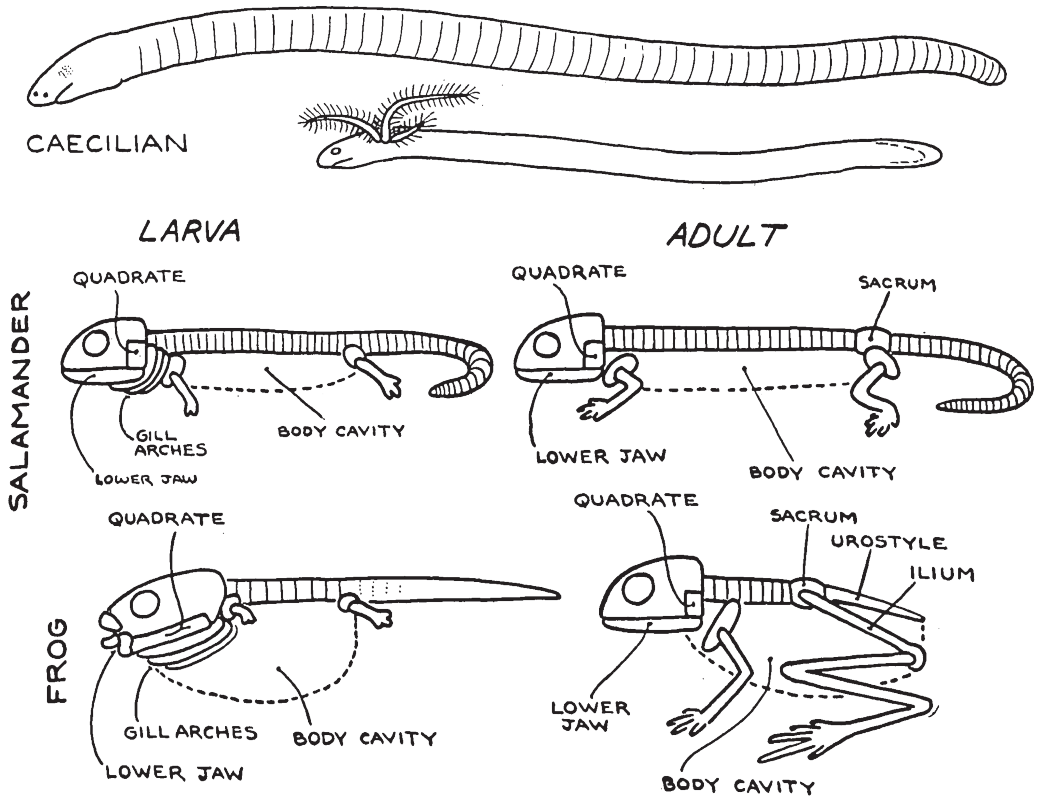


Fig. 1.1. Living amphibians: Gymnophiona (caecilians), Caudata (salamanders) and Anura (frogs). Sketches of adult and larval stages showing some anatomical structures. Adapted from Orton (1953).

Living frogs display a great diversity in their external appearance, life history, ecology and behaviour and yet have retained a highly conserved body plan for more than 200 million years (Handrigan and Wassersug, 2007). Despite the high diversity in cranial architecture, the typical anuran skull is broad and fenestrated, with a reduced number of cranial elements (relative to other amphibians), the jaw articulation is located toward the posterior limit of the skull, the dentition is reduced, and the cranial elements involved in sensory systems (*e.g.,* olfaction, hearing) are much more elaborated than those of salamanders and caecilians (Duellman and Trueb, 1986).

The axial skeleton is also highly modified compared to that of salamanders and caecilians (Fig. 1.1). There is a great variation in vertebrae centrum structure, a feature that attracted the attention of earlier workers such as Cope (1865) and Noble (1922), who used it in their classifications (Duellman and Trueb, 1986). Compared to other extant and fossil amphibians (*e.g., Vieraella*; Fig. 1.2), living frogs possess a shortened vertebrate column with 5-9 presacral vertebrae (Fig. 1.1), which posses transverse processes (except in the atlas) and are firmly articulated allowing only moderate lateral and dorsoventral flexure (Duellman and Trueb, 1986). Ribs are generally absent, with the exception of *Triadobatrachus* and *Vieraella* and some basal living families: ribs are free in *Leiopelma*, *Ascaphus* and Discoglossoidea, and ankylosed to the transverse processes in adult

Pipidae (Duellman and Trueb, 1986). Frogs possess a single sacrum in which the pelvic girdle is suspended, and the postsacral vertebrae are fused into a single rodlike structure (urostyle), which lies between the shafts of the ilia of the pelvic girdle and bears muscular attachments to these elements (Fig. 1.1) (Duellman and Trueb, 1986).

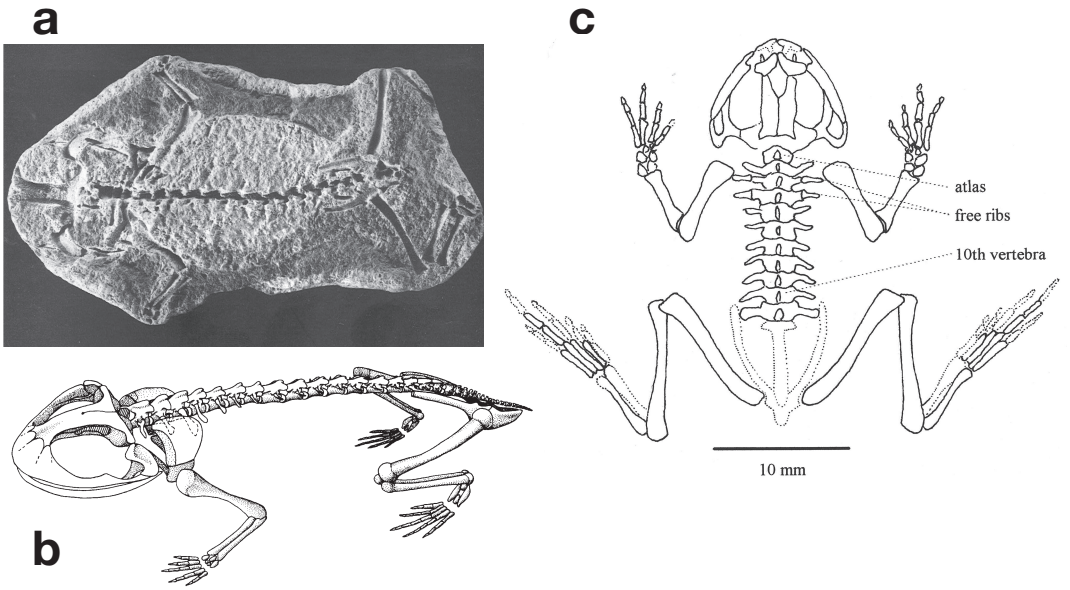


Fig. 1.2. Fossil Salientians. (a) Fossil inprint (dorsal side) of *Triadobatrachus massinoti*; Modified from Roček and Rage (2000). (b) Hypothesized reconstruction of the skeleton of *Triadobatrachus massinoti*. Total length about 10 cm. Modified from Rage and Roček (1998). (c) Restoration of the skeleton of *Vieraella herbti*. Modified from Roček (2000).

One of the most distinctive features of frogs are their elongated hind limbs and feet for saltatory locomotion, a unique feature among tetrapods representing a fundamental departure from the generalized mode of progression by alternating limb movement (Duellman and Trueb, 1986). Many of the morphological features of anurans are related to this mode of locomotion, including the short tailless body and long hind limbs, as well as a highly modified pelvic girdle and powerful musculature (Duellman and Trueb, 1986). Many of them were already present in *Triadobatrachus* and *Vieraella* (Fig. 1.2), suggesting that saltatory locomotion had already evolved in the Triassic (Carroll, 2009). The high structural diversity of the pelvic girdle and associated musculature (specially that of the thigh) has traditionally been used for taxonomic purposes (Noble, 1922). Pectoral girdle structure has also been widely used in taxonomy, two basic forms being defined based on whether the epicoracoid cartilages are composed of two overlapping plates (arciferal), or are fused along the ventral midline (firmistermal) (Emerson, 1984), although intermediate forms (pseudoarciferal or pseudofirmistermal) have also evolved secondarily from firmistermal and arciferal girdles, respectively (Duellman and Trueb, 1986). The arcifery condition is the most common among frogs, and it is thought to represent the ancestral condition, from which firmistermy evolved several independent times (Emerson, 1984).

1.2. The phylogeny of frogs: history, current knowledge and controversies

The monophyly of Anura has generally not been questioned, being supported by a large body of literature on both morphological (*e.g.,* Rage and Janvier, 1982; Milner, 1988; Benton, 1990; Trueb and Cloutier, 1991; Haas, 2003; Marjanović and Laurin, 2007; Ruta and Coates, 2007) and molecular data (Hay et al., 1995; Feller and Hedges, 1998; Roelants and Bossuyt, 2005; San Mauro et al., 2005; Zhang et al., 2005a; Roelants et al., 2007). As an exception, Roček and Vesely (1989) suggested a diphyletic origin of Anura based on the presumable non-homology of a larval character, but the homology of these structures was latter firmly established (Olsson and Hanken, 1996; de Sá and Swart, 1999).

For many years, the understanding of the phylogenetic relationships among frogs was based on relatively small number of morphological characters. Traditional classifications used few characters, mainly the number or vertebrae and shape of their centrum, pectoral girdle, ribs, thigh musculature and dentition (see Fig. 1.3) (Cope, 1865; Boulenguer, 1882; Noble, 1922, 1931). Inger (1967) provided the first quantitative study of frog phylogeny based on 10 morphological characters for 12 families, and proposed three alternative phylogenies. Kluge and Farris (1969) introduced the Wagner parsimony and at the same time reanalyzed the dataset of Inger (1967), producing a fourth alternative phylogeny. Lynch (1973) divided the families of frogs into "archaic", "transitional", and "advanced", reflecting three "evolutionary grades" defined by the number of primitive characters retained by each family. Starrett (1973) proposed a completely different scenario for anuran evolution based on the four larval types defined by Orton (1953, 1957), which were originally devised as a phenetic (not phylogenetic) classification system (Duellman and Trueb, 1986). Sokol (1975) reinterpreted the data of Starrett (1973), obtaining a phylogeny much more congruent with previous studies (Kluge and Farris, 1969; Lynch, 1973). Duellman (1975) incorporated all this previous knowledge into a formal classification of Anura, in which two major suborders were recognized: Archaeobatrachia ("archaic frogs") and Neobatrachia ("advanced frogs"), using the names proposed by Reig (1958). Laurent (1979; 1985) recognized an additional suborder, Mesobatrachia, including some families subtracted from Duellman's Archaeobatrachia. Duellman and Trueb (1986) presented a phylogeny based on 16 morphological characters for 21 families, and recovered five major lineages of frogs (Fig. 1.4).

In phylogenetic analyses of frogs using morphological characters, the polarization of character states proved to be difficult due to the big morphological gap between all three groups of amphibians and the scarcity of intermediate fossils (Kluge and Farris, 1969). This, along with

the small number of available characters, produced weakly supported hypotheses that were excessively influenced by the interpretations of different authors about which characters would more reliably reflect the course of evolution (Kluge and Farris, 1969; Frost et al., 2006). In addition, the above phylogenetic analyses (i) used families as terminal taxa (rather than species), producing an excessive generalization of characters and (ii) assumed the monophyly of every family, despite the evidence of paraphyly for at least some of them (Ford and Cannatella, 1993; Frost et al., 2006). Overall, phylogenetic relationships of frogs long remained a field of conflicting opinion, isolated lines of evidence and unsubstantiated assertion of paraphyly and polyphyly (Frost et al., 2006).

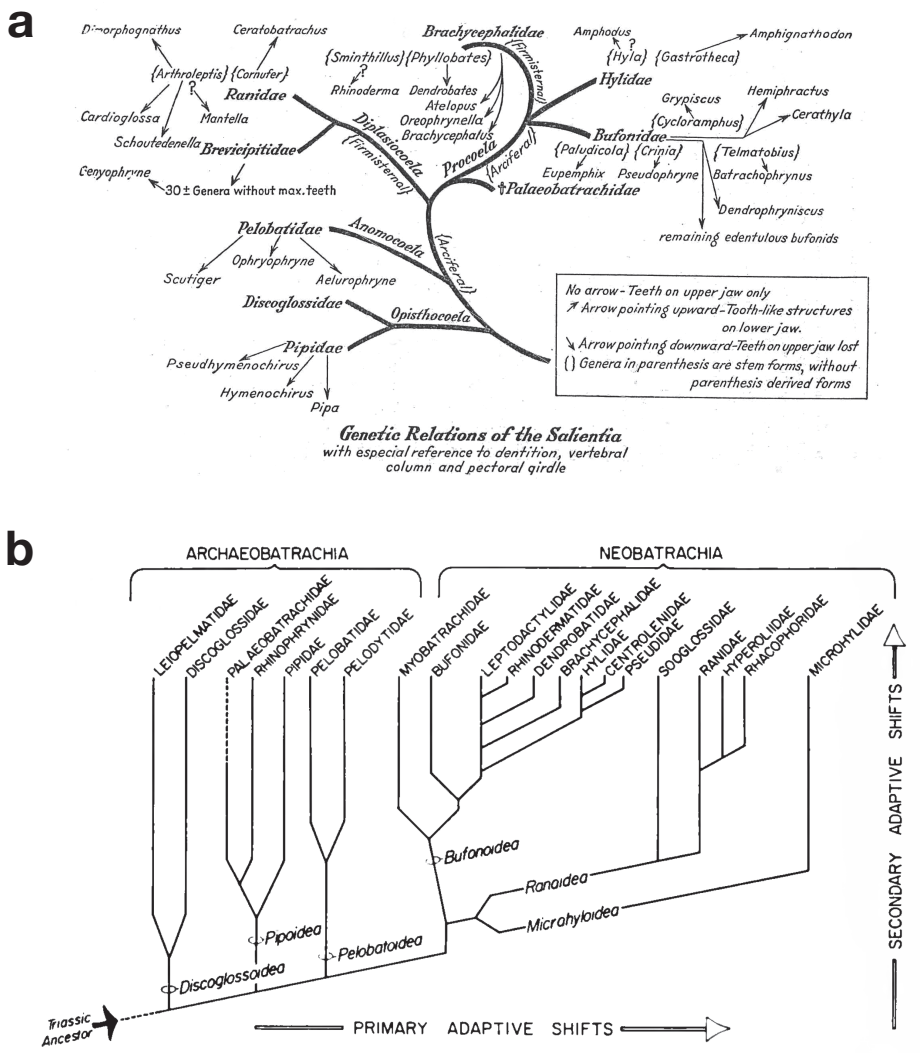


Fig. 1.3. Phylogenetic relationships of Anura according to (a) Noble (1922) and (b) Duellman (1975). Note that the classifications by Noble and Duellman do not directly reflect their evolutionary sketches, as both authors maintained paraphyletic groups in their classifications.

First phylogenetic analyses using larger number of characters (124-180) and species as terminal taxa (41-42 taxa) were those performed by Cannatella (1985) and Ford (1989), their results being synthesized into a review paper in which they proposed an unranked classification of Anura based strictly on branching patterns (Ford and Cannatella, 1993). The phylogeny of frogs presented by Ford and Cannatella (1993) supported a monophyletic Neobatrachia and Mesobatrachia (*sensu* Laurent, 1979), but pointed out to a paraphyletic origin of "Archaeobatrachia" (*sensu* Duellman, 1975) (Fig. 1.4). First molecular studies based on partial sequences of mitochondrial (mt) rRNA genes (Hedges and Maxson, 1993; Hay et al., 1995; Dutta et al., 2004) found Archaeobatrachia and Neobatrachia to be reciprocally monophyletic (Fig. 1.4), following the dichotomy of traditional classifications (Duellman, 1975). In contrast, an analysis of nuclear rRNA genes (Hillis et al., 1993) rejected the monophyly of both "Archaeobatrachia" and "Neobatrachia", but not that of Mesobatrachia. A combination of the these nuclear rRNA and morphological data (from Duellman and Trueb, 1986) supported a monophyletic Neobatrachia and Mesobatrachia, and rejected the monophyly of "Archaeobatrachia" (Hillis et al., 1993). A second study of nuclear rRNA genes taking into account secondary structure in the analyses supported a monophyletic Neobatrachia, but rejected the monophyly of both "Archaeobatrachia" and "Mesobatrachia" (Kjer, 1995).

In recent years, the use of larger molecular and morphological datasets, along with the development of sophisticated methods for probabilistic phylogenetic inference allowed resolving some of the controversial issues regarding higher-level relationships among living frogs. Yet, many important questions still remain under debate. Most recent morphological and molecular studies using larger datasets, were congruent in supporting a monophyletic origin for Neobatrachia (*e.g.*, Biju and Bossuyt, 2003; San Mauro et al., 2004a; Pyron and Wiens, 2001), and the paraphyletic nature of both "Archaeobatrachia" and "Mesobatrachia" (San Mauro et al., 2004a; Roelants and Bossuyt, 2005; San Mauro et al., 2005; Roelants et al., 2007; Pyron and Wiens, 2011). The above studies generally recognize four major lineages among living anurans, (i) Discoglossoidea or Costata, (ii) Pipoidea or Xenoanura, (iii) Pelobatoidea or Anomocoela, and (iv) Neobatrachia, as well as two basal genera of uncertain placement (*Ascaphus* and *Leiopelma*) sometimes grouped under the name Amphicoela.

Table 1.1 shows the distribution of families into these major lineages of frogs, along with the number of species included in each group. Most of the above studies have supported a sister group relationship between Neobatrachia and Pelobatoidea (San Mauro et al., 2004a; Roelants and Bossuyt, 2005; San Mauro et al., 2005; Roelants et al., 2007). In addition, Discoglossoidea and Pipoidea are generally recovered as lineages that branch successively after the basal genera *Ascaphus* and *Leiopelma*, although the branching order varied between studies: Discoglossoidea branching before (Lynch, 1973; Duellman and Trueb, 1986; Haas, 1997; Roelants and Bossuyt, 2005; Roelants et al., 2007) or after Pipoidea (Haas, 2003; San Mauro et al., 2005; Frost et al., 2006) (Fig. 1.4). However, some studies found alternative phylogenetic hypotheses of sister-

group relationships between Discoglossoidea and Pipoidea (Biju and Bossuyt, 2003; Gissi et al., 2006), Pipoidea and Pelobatoidea ("Mesobatrachia"; García-París et al., 2003b), or Pipoidea and Neobatrachia (Hoegg et al., 2004). Some studies based on morphology (mostly larval) have also found a basal position of Pipoidea, as sister group to all other frogs (Fig. 1.4) (Maglia et al., 2001; Púgener et al., 2003). Overall, two main questions remain contentious regarding phylogenetic relationships among major lineages of frogs: (i) the relative phylogenetic position of the basal genera *Ascaphus* and *Leiopelma*, and (ii) the relationship between Discoglossoidea and Pipoidea.

Table 1.1. The major lineages of living frogs, and their families (with the number of species). The families shown follow the recent taxonomy of Amphibian Species of the World (Frost, 2011).

Family	No. Species	Family (Cont.)	No. Species (Cont.)
AMPHICOELA	**6**	Craugastoridae	115
Leiopelmatidae	6	Cycloramphidae	105
DISCOGLOSSOIDEA or COSTATA	**20**	Dendrobatidae	179
Alytidae	12	Eleutherodactylidae	201
Bombinatoridae	8	Hemiphractidae	94
PIPOIDEA or XENOANURA	**34**	Hylidae	901
Pipidae	33	Hylodidae	42
Rhinophrynidae	1	Leiuperidae	86
PELOBATOIDEA or ANOMOCOELA	**170**	Leptodactylidae	100
Megophryidae	156	Strabomantidae	569
Pelobatidae	4	**Afrobatrachia**	**389**
Pelodytidae	3	Arthroleptidae	140
Scaphiopodidae	7	Brevicipitidae	31
NEOBATRACHIA	**5736**	Hemisotidae	9
Calyptocephalellidae	4	Hyperoliidae	209
Heleophrynidae	7	**Microhyloidea**	**487**
Myobatrachoidea	**128**	Microhylidae	487
Limnodynastidae	43	**Natatanura**	**1374**
Myobatrachidae	85	Ceratobatrachidae	85
Sooglossoidea	**5**	Dicroglossidae	171
Nasikabatrachidae	1	Mantellidae	191
Sooglossidae	4	Micrixalidae	11
Nobleobatrachia	**3342**	Nyctibatrachidae	17
Allophrynidae	1	Petropedetidae	18
Aromobatidae	104	Phrynobatrachidae	82
Bufonidae	558	Ptychadenidae	53
Brachycephalidae	51	Pyxicephalidae	68
Centrolenidae	146	Ranidae	347
Ceratophryidae	86	Ranixalidae	10
Ceuthomantidae	4	Rhacophoridae	321

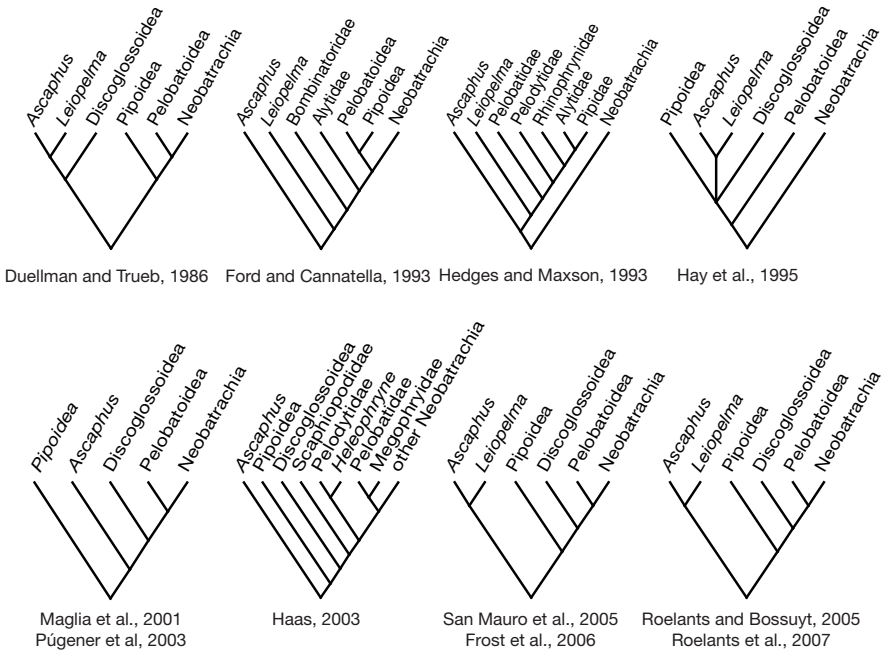


Fig. 1.4. Alternative phylogenetic hypotheses proposed for the major lineages of frogs (with the corresponding references). In Maglia et al. (2001) the relationship between Bombinatoridae and Alytidae is not resolved (it is a polytomy), and in Púgener et al. (2003) Neobatrachia is paraphyletic because *Hyla* appears more closely related to Pelobatoidea than to other neobatrachians.

1.2.1. The root of the anuran tree of life

Two genera with disjoint distributions have been traditionally considered the most basal among extant frogs: *Leiopelma*, with four extant and three extinct species, which are endemic to New Zealand, and *Ascaphus*, with two species living in North America (Fig. 1.5) (Duellman and Trueb, 1986; Green and Cannatella, 1993). *Leiopelma* and *Ascaphus* represent key lineages in understanding anuran evolution, but their exact phylogenetic position is rather elusive, and several alternative hypotheses have been proposed.

Both *Ascaphus* and *Leiopelma* share several morphological characters (Green et al., 1989; Green and Cannatella, 1993) that were often used to hypothesize their sister group relationship. The most important one is the presence of amphicoelous vertebrae (*i.e.,* with concavities on both anterior and posterior ends), which led them to be grouped under the name Amphicoela (Noble, 1922, 1931; Ritland, 1955; Green and Cannatella, 1993). Additionally, *Ascaphus* and *Leiopelma* share the presence of nine presacral vertebrae, paired caudalipuboischiotibialis (tail-wagging) muscles, an epipubic cartilage, ribs not fused to the vertebrae, absence of vocal sacs, and the absence of a columella (Green and Cannatella, 1993). Most of these characters are now considered symplesiomorphies retained by both groups, and therefore not suitable for

phylogenetic inference (Hennig, 1966). The secondary loss of the columella may be the only true character with synapomorphic state (Stephenson, 1951). Ritland (1955) pointed out that the tail-wagging muscles are likely non-homologous to those of salamanders, and hence they may represent another synapomorphy for *Ascaphus* and *Leiopelma*. Moreover, each genus exhibits its own unique apomorphies: *Ascaphus* has an intromittent organ for copulation in males (Fig. 1.5) (Duellman and Trueb, 1986) and a highly modified torrent-dwelling tadpole, whereas *Leiopelma* has ventral inscriptional ribs (Noble, 1931; Laurent, 1986; Ford and Cannatella, 1993) and lacks a feeding larval stage (Archey, 1922; Altig and Johnston, 1989; Bell and Wassersug, 2003).

Even though *Ascaphus* and *Leiopelma* have always been considered basal living frogs, the possession of many symplesiomorphic characters along with their own autapomorphies have at the same time hindered their specific phylogenetic relationships with respect to all other anuran lineages. Based on morphological evidence, the following groups have been proposed alternatively as the most basal lineage of frogs, *i.e.,* being sister group to all other anurans: (i) *Leiopelma* + *Ascaphus* ("Amphicoela" hypothesis; Lynch, 1973); (ii) *Leiopelma* + *Ascaphus* + Discoglossoidea (including *Bombina*, *Barbourula*, *Discoglossus*, and *Alytes*) (Duellman and Trueb, 1986); (iii) *Ascaphus* alone (Ford and Cannatella, 1993); and (iv) Pipoidea (Maglia et al., 2001; Púgener et al., 2003). Ford and Cannatella (1993) proposed five putative synapomorphies to support *Ascaphus* as the sister group to the clade named Leiopelmatanura (*Leiopelma* + all other anurans). Recent molecular studies favoured the first hypothesis suggested by Lynch (1973) (Roelants and Bossuyt, 2005; San Mauro et al., 2005; Frost et al., 2006; Roelants et al., 2007; Pyron and Wiens, 2011).

1.2.2. Discoglossoidea

The oldest discoglossoid fossil known is *Eodiscoglossus* from the Middle Jurassic of Europe (Evans et al., 1990; Martín and Sanchíz, 2010). The monophyly of the group is strongly supported (*e.g.,* San Mauro et al., 2004a), in spite of Ford and Cannatella (1993), who recovered *Ascaphus* and *Leiopelma* as most basal lineages, and the family Bombinatoridae as sister group to a clade named Discoglossanura, which included the family Alytidae and all other remaining frogs. The Discoglossoidea include four living genera (Fig. 1.5) *Alytes, Bombina, Barbourula,* and *Discoglossus*, traditionally grouped into a single family (Discoglossidae; Duellman, 1975; Duellman and Trueb, 1986), but latter divided into Bombinatoridae (*Bombina* and *Barbourula*) and Alytidae [=Discoglossidae] (*Alytes* and *Discoglossus*) (San Mauro et al., 2004a). *Bombina* and *Barbourula* have generally been considered sister taxa, but their phylogenetic affinities to the other two discoglossoidean genera have been disputed (San Mauro et al., 2004a). Most studies supported a sister-group relationship between *Alytes* and *Discoglossus* to the exclusion of *Bombina* (San Mauro et al., 2004a; Frost et al., 2006; Blackburn et al., 2010; Pyron and Wiens, 2011), even though a sister-group relationship between *Bombina* and *Alytes* (Lanza et al., 1975) or between *Bombina* and *Discoglossus* (Maxson and Szymura, 1984; Haas, 2003) have also been suggested.

DISCOGLOSSOIDEA

Leiopelma archeyi
© D. M. Green, 2004

Ascaphus truei
© B. Moon, 2000

Alytes cisternasii
© I. Martínez-Solano, 2006

Bombina orientalis
© B. Akeret, 2009

PIPOIDEA

Pipa pipa
© I. de la Riva, 2003

Rhinophrynus dorsalis
© D. B. Fenolio, 2007

PELOBATOIDEA

Pelobates cultripes
© I. Martínez-Solano, 2006

Megophrys nasuta
© G. M. Rosa, 2009

Pseudhymenochirus merlini
© F. Glaw, 2010

Xenopus fraseri
© P. Janzen, 2009

Scaphiopus couchii
© B. Moon, 2003

Pelodytes punctatus
© I. Martínez-Solano, 2006

NEOBATRACHIA

Eleutherodactylus coqui
© R. I. Márquez, 2004

Leptodactylus fallax
© G. M. Rosa, 2009

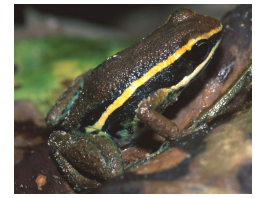
Bufo bufo
© I. Martínez-Solano, 2006

Dyscophus antongilii
© G. M. Rosa, 2009

Heleophryne regis
© M. Vences, 1998

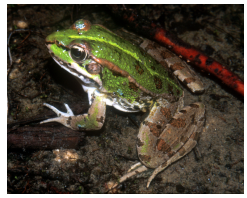
Boophis bottae
© G. M. Rosa, 2009

Hypsiboas albopunctatus
© I. de la Riva, 2003



Ameerega boliviana
© I. de la Riva, 2003

Sooglossus thomasseti
© N. Doak



Rana perezi
© I. Martínez-Solano, 2006

Telmatobius gigas
© J. M. Padial, 2009

Rhacophorus prominanus
© G. M. Rosa, 2009

Fig. 1.5. Anuran diversity across all major lineages.
Pictures from Amphibiaweb (http://amphibiaweb.org/) and Arkive (http://www.arkive.org/).

1.2.3. Pipoidea

The first known pipoid fossil, *Rhadinosteus*, dates back from the Late Jurassic–Early Cretaceous (Henrici, 1998; Martín and Sanchíz, 2010). The clade Pipoidea is well supported (Maglia et al., 2001; Haas, 2003; Púgener et al., 2003; San Mauro et al., 2005; Frost et al., 2006), and includes the living families Pipidae and Rhinophrynidae (whose only living representative is *Rhinophrynus dorsalis*; Fig. 1.5), as well as the fossil family Palaeobatrachidae (Špinar, 1972). The family Pipidae was firmly established by Noble (1922), and it was further supported by many synapomorphies in posterior studies (Cannatella and Trueb, 1988a; Ford and Cannatella, 1993; Haas, 2003). The extant members of this family include the South American genus *Pipa* (Surinam toads) and the four African genera *Hymenochirus, Silurana, Xenopus, and Pseudhymenochirus* (African clawed frogs) (Fig. 1.5). The monotypic genus *Pseudhymenochirus* has been poorly studied in the past due to its rarity in collections (Cannatella and Trueb, 1988b). Initially, it was grouped with *Hymenochirus* (Chabanaud, 1921), but later regarded as "intermediate" between *Hymenochirus* and *Xenopus* (Noble, 1931; Dunn, 1948; Báez and Harrison, 2005) or considered a "primitive" *Hymenochirus* (Sokol, 1977). At present, a sister group relationship between *Pseudhymenochirus* and *Hymenochirus* is widely accepted (forming the Hymenochirini; Cannatella and Trueb, 1988b). Geographically, *Pseudhymenochirus* merlini is separated by 2000 Km from the westernmost *Hymenochirus* species in Nigeria (Menzies, 1967).

Morphologically, pipids are rather aberrant compared to all other frogs (Fig. 1.5; Cannatella and Trueb, 1988a; Cannatella and de Sá, 1993). Initially, this morphology was considered to be relatively ancestral among frogs, and many of their characters to retain plesiomorphic states. In contrast, pipids are now viewed as highly derived frogs with many autapomorphies primarily related to their fully aquatic lifestyle (Cannatella and de Sá, 1993). Pipids represent a nice example of highly adapted form and function that evolved from an inherited frog bauplan, which is *per se* highly specialized within amphibians (and tetrapods), and restricted to limited variation (Emerson, 1988). Pipids are the only fully aquatic group of frogs, and their derived morphology and biology are largely a product of adaptations to this lifestyle (Cannatella and Trueb, 1988a).

In several respects pipids have been more extensively studied than any other group of frogs because *Xenopus laevis* and *Silurana tropicalis* (grouped under the subfamily Dactylethrinae; Noble, 1931) are considered model systems in physiology, development, and cell and molecular biology (*e.g.,* Cannatella and de Sá, 1993). Knowledge on the closest relatives of model organisms is crucial to interpret and understand the evolutionary origin of studied characters and functions, but remarkably the phylogenetic relationships of pipids have not been comprehensively assessed so far.

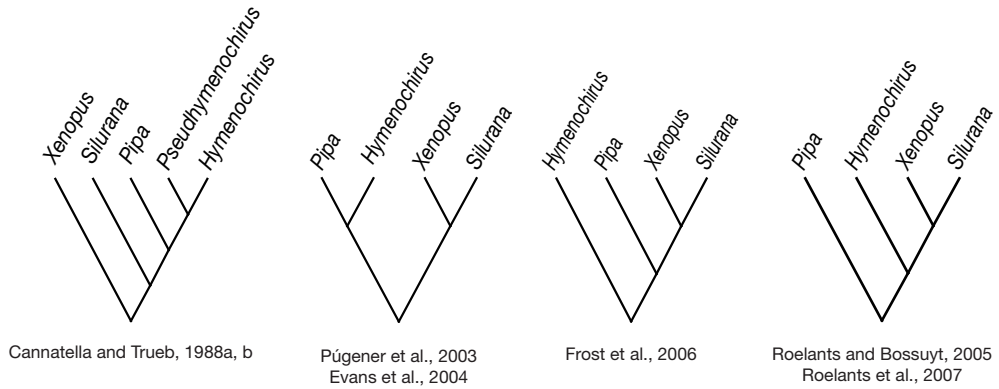


Fig. 1.6. Alterantive phylogenetic hypotheses proposed for the genera of the family Pipidae.

The phylogenetic relationships within Pipidae remain controversial, and almost all possible alternative hypotheses have been proposed for the relationships among the five recognized genera (Fig. 1.6): (i) (*Xenopus* + (*Silurana* + (*Pipa* + (*Hymenochirus* + *Pseudhymenochirus*)))) (Cannatella and Trueb, 1988a, b); (ii) ((*Pipa* + *Hymenochirus*) + (*Xenopus* + *Silurana*)) (Maglia et al., 2001; Púgener et al., 2003; Evans et al., 2004; Evans et al., 2005a; Trueb et al., 2005); (iii) (*Hymenochirus* + (*Pipa* + (*Xenopus* + *Silurana*))) (Frost et al., 2006); (iv) (*Pipa* + (*Hymenochirus* + (*Xenopus* + *Silurana*))) (Roelants and Bossuyt, 2005; Roelants et al., 2007; Pyron and Wiens, 2011). The latter hypothesis is also consistent with other studies with a smaller taxon sampling (de Sá and Hillis, 1990; San Mauro et al., 2005). Therefore, phylogenetic relationships among pipid genera require further study and clarification.

1.2.4. Pelobatoidea

Most known fossils of pelobatoids date back to the Cenozoic (Martín and Sanchíz, 2010), even though the first record might be represented by a pelodytid from the Late Jurassic of North America (Martín and Sanchíz, 2010). The Pelobatoidea is generally acknowledged to be monophyletic in most recent studies (*e.g.,* Ford and Cannatella, 1993; García-París et al., 2003b; Púgener et al., 2003; Roelants and Bossuyt, 2005; San Mauro et al., 2005; Frost et al., 2006) but it was not in the past (*e.g.,* Lynch, 1973). In addition, two recent morphological analysis using larval characters found Pelobatoidea to be paraphyletic, because either *Heleophryne* (Haas, 2003) or *Hyla* (Púgener et al., 2003) were grouped with pelobatoids.

Currently, four families are recognized: Pelobatidae (genus *Pelobates*), Scaphiopodidae (genera *Scaphiopus* and *Spea*) Pelodytidae (genus *Pelodytes*), and Megophryidae (the most diversified family including 10 genera and about 156 species) (Table 1.1. and Fig. 1.5). Traditionally, the family Scaphiopodidae was included into Pelobatidae (Duellman and Trueb, 1986), until both morphological (Haas, 2003) and molecular studies (García-París et al., 2003b) showed that they represent

independent lineages. The monophyly of all other families has generally not been questioned (Ford and Cannatella, 1993). Phylogenetic relationships among pelobatoid families have long remained controversial and no fewer than 12 hypotheses have been proposed (Wiens and Titus, 1991; Henrici, 1994; Barbadillo et al., 1997; Lathrop, 1997; Gao and Wang, 2001). Currently, the most widely accepted hypothesis is Scaphiopodidae + (Pelodytidae + (Pelobatidae + Megophryidae) (García-París et al., 2003b), although Frost et al. (2006) found sister group relationships between Pelodytidae and Scaphiopodidae, and between Pelobatidae and Megophryidae.

1.2.5. Neobatrachia

The fossil evidence for Neobatrachia greatly underestimates the probable date of origin of this group, as the earliest known fossil is a calyptocephalellid from the Eocene (Báez, 2000; Martín and Sanchíz, 2010), but Neobatrachia already separated from Pelobatoidea in the Late Triassic (*e.g.,* Roelants et al., 2007). Neobatrachia is an evolutionarily highly successful clade that contains more than 96% of the overall species diversity of living amphibians (Table 1.1 and Fig. 1.5; Frost, 2011). The monophyly of Neobatrachia has generally not been questioned, although two morphological analyses found it to be paraphyletic, because *Heleophryne* was grouped with *Pelodytes* away from all other neobatrachians (Haas, 2003), and because *Hyla* was grouped with pelobatoideans (Púgener et al., 2003), rendering Pelobatoidea as paraphyletic as well. Traditionally, neobatrachian frogs have been divided into "Hyloidea" and "Ranoidea" according to the shape of the vertebral centrum, pectoral girdle architecture, and conformation of thigh musculature (Lynch, 1973; Ford and Cannatella, 1993; Darst and Cannatella, 2004). However, these groups are paraphyletic as traditionally defined (Ford, 1989; Biju and Bossuyt, 2003; van der Meijden et al., 2007a), and thus they have been redefined including a more restricted number of families. Currently, two major clades are recognized within Neobatrachia: Ranoides, which comprises three well-supported monophyletic groups (Afrobatrachia, Microhyloidea, and Natatanura), and Hyloides, including a well-supported monophyletic Nobleobatrachia (Frost et al., 2006; Bossuyt and Roelants, 2009). However, the phylogenetic position of basal families such as Heleophrynidae, Sooglossidae, Nasikabatrachidae, Limnodynastidae, Calyptocephalellidae, or Myobatrachidae remain unresolved (Frost et al., 2006).

Previous studies have shown that neobatrachian frogs exhibit higher mt substitution rates compared to their non-neobatrachian relatives (Hay et al., 1995; Feller and Hedges, 1998; Hoegg et al., 2004; San Mauro et al., 2004a; Gissi et al., 2006). Yet, it is not clear when the shifts in substitution rates precisely occurred. Moreover, a previous study showed that some nuclear genes also possess higher substitution rates in neobatrachians (Hoegg et al., 2004), but it is unknown how general this trend is. On the one hand, the heterogeneous distribution of substitution rates among lineages of frogs, together with the use of genetically distant outgroups (the closest living sister taxa of frogs are salamanders; Zardoya and Meyer, 2001) has been shown to be

the source of several phylogenetic artefacts, such as monophyly of non-neobatrachian frogs ("Archaeobatrachia": Hedges and Maxson, 1993; Hay et al., 1995) or the incorrect phylogenetic placement of Neobatrachia due to long-branch attraction effects in different parts of the frog tree (Gissi et al., 2006). The unequal distribution of mt substitution rates across the anuran tree has also been suggested to yield considerably older time estimates for divergences among neobatrachians (Igawa et al., 2008). Moreover, it has been suggested that the shift in mt substitution rates in Neobatrachia could be related with the higher diversification rates observed in Nobleobatrachia and Ranoides (Hoegg et al., 2004). To answer all these open questions, it is necessary to precisely delimit the node in the anuran phylogeny where the shift in evolutionary rates took place.

1.2.6. The active field of amphibian taxonomy and systematics

The taxonomy and systematics of anurans (and that of amphibians in general) is in a state of rapid growth and change due to the discovery and description of a increasing number of new species (Vitt and Caldwell, 2009). In the 1980s, Duellman and Trueb (1986) recognized 3,438 species of frogs, divided into 21 families and 301 genera, in contrast to the 5,966 species, 49 families and 407 genera currently accepted (Frost, 2011). This increase in the number of species has been in part the product of rigorous taxonomic and systematic studies of known populations of described species that have been found to be genetically or bioacustically distinct. As a consequence, many large genera, families or higher taxa have been examined in detail (specially during the last decade), and are being appropriately systematized and divided into monophyletic groups (Vitt and Caldwell, 2009). However, the removal from synonymy represents only a small proportion of the newly described species (approx. 14%; Stuart et al., 2008) and the vast majority correspond to genuine discoveries, as the relatively recent finding of the new family Nasikabatrachidae (Biju and Bossuyt, 2003) or the number of new frog species from Madagascar awaiting formal description (Köhler et al., 2005; Vieites et al., 2009). Currently, largely unexplored territories, mainly South America and Southeast Asia, are the source of most newly described species, the top four countries being Brazil, Papua New Guinea, Sri Lanka, and Peru (Amphibiaweb, 2011). Unfortunately, at the same time, the declining of many populations due to anthropogenic pressures and the spread of diseases is producing a significant increase in the extinction of frog species (Amphibiaweb, 2011).

Several studies boosted the field of frog systematics and moved toward a taxonomy consistent with phylogenetic relationships (*e.g.,* Faivovich et al., 2005; San Mauro et al., 2005; Grant et al., 2006; Roelants et al., 2007; Pyron and Wiens, 2011). "The Amphibian Tree of Life" (Frost et al., 2006) is the first large-scale project on amphibian systematics and taxonomy and supposed a landmark in the field. Currently, amphibian taxonomy is being continuously updated by Amphibiaweb (Amphibiaweb, 2011) and Amphibian Species of the World (Frost, 2011), whose contribution is invaluable.

1.3. Ecological, behavioural, and phylogenetic diversity, with emphasis on sound production

1.3.1. Life history

The life histories of anurans are highly diversified. Fertilization is generally external, and the generalized (and presumably primitive) mode of life history involves a biphasic cycle with aquatic eggs and larvae (*i.e.,* indirect development) (Duellman and Trueb, 1986). The tadpole is structurally, physiologically, ecologically, and behaviourally different from the fully developed adult and both phases are joined by a major metamorphic event, which requires a major reorganization of anatomy and physiology (Vitt and Caldwell, 2009).

A departure from this generalized mode is represented by species with internal fertilization. There are few examples (about 20), being remarkable the case of *Ascaphus*, whose adult males posses an intromittent organ for fertilization (Townsend et al., 1981; Vitt and Caldwell, 2009). Remarkably, anurans display a diversity of reproductive modes, classified into more than 29 general types (Duellman and Trueb, 1986). Eggs can be directly deposited into water bodies (ponds, streams, etc.), land, or vegetation (tadpoles may even develop into axils of plants filled with water); or they can be carried by the adult (Duellman and Trueb, 1986). Some species construct foam nests to avoid desiccation in either water, land or vegetation; and tadpoles can develop either inside the foam nests (in the case of non-feeding tadpoles that possess a high yolk content), or into water after post-hatching tadpoles drop from nests constructed above water bodies or after being washed by the rain (Duellman and Trueb, 1986). Eggs may be carried by the male or the female in the dorsum or in dorsal pouches; for a short period (tadpoles then develop into water) or until development is complete (Duellman and Trueb, 1986). Few species (about five; Wake, 1978) retain eggs into oviducts; some have non-feeding tadpoles that nourish from the yolk (*i.e.,* they are ovoviviparous), whereas others produce oviductal secretions for the developing young (*i.e.,* they are viviparous) (Duellman and Trueb, 1986). In certain species, the tadpole stage is absent and eggs produce small froglets (*i.e.,* development is direct) (Lynch, 1971; Bell, 1978). Parental care may also be present, either in the mode of attendance of eggs, transportation of egg or larvae, or feeding of tadpoles (Duellman and Trueb, 1986). *Pipa* is a remarkable example of a frog with direct development and parental care, whose females carry the eggs into the back and are covered by a special vascular tissue; eggs develop into tadpoles embedded into females' back (*P. carvalhoi*), or directly into froglets (*P. pipa*) (Rabb, 1960; Vitt and Caldwell, 2009).

This great diversity of reproductive modes is unequally distributed across the phylogeny: some clades are characterized by having more than 10 different reproductive modes (*e.g.,* Myobatrachoidea), whereas whole families are characterized by possessing a single mode (*e.g.,* Pelodytidae or Dendrobatidae) (Duellman and Trueb, 1986). The evolution of reproductive strategies in anurans has been proposed to follow a trend towards terrestriality: the production of eggs with sufficient yolk and non-feeding tadpoles may represent preadaptations for terrestriality, whereas direct development may have evolved from non-feeding tadpoles (Duellman and Trueb, 1986). For example, within egg-brooding hylid frogs, there is a sequence of specialization from hatching tadpoles, to non-feeding tadpoles and direct development into froglets (Duellman and Maness, 1980). Direct development is known to have evolved independently at least 15 times across anuran phylogeny (Hanken et al., 1997).

Tadpoles are post-hatching larval stages that are generally aquatic and obtain nutrients from the environment to further develop and grow (Duellman and Trueb, 1986). The larvae and the adults are subjected to very different selective pressures, and exploit different sets of resources, generally avoiding competition for food or shelter (Duellman and Trueb, 1986). Tadpoles possess a series of adaptations to their particular lifestyle, mostly related to respiration and feeding. They have internal gills and specialized mouthparts for feeding (they lack true teeth), as well as specialized branchial apparatus for both buccal pumping and feeding (Duellman and Trueb, 1986). Many of the structural differences among larvae are associated with their biology rather than with phylogenetic inheritance (Duellman and Trueb, 1986). Nevertheless, larval characters have extensively been used to infer phylogenetic relationships among frogs (*e.g.,* Starrett, 1973; Maglia et al., 2001; Haas, 2003; Roelants et al., 2011). According to Púgener et al. (2003), larval characters might provide a more suitable set of characters than those of the adult, because they are very derived in basal anurans. Tadpoles have been classified into four types based on the structure of the opercular chamber and its opening(s) from the body, and the structure of the larval mouth (Orton, 1953). Type I (Xenoanura) is typical of Pipidae and Rhinophrynidae, type II (Scoptanura) is present in Microhylidae, type III (Lemmanura) occurs in Leiopelmatidae and Alytidae, and type IV (Acosmanura) includes all other frogs (Orton, 1953; Haas, 2003).

In most species of frogs, aquatic larval stages undergo a series of abrupt postembryonic changes involving structure, physiological, biochemical, and behavioural transformations, a process known as metamorphosis (Duellman and Trueb, 1986). One example of profound modifications is represented by the branchial apparatus, which is related to respiration and feeding in tadpoles, but later forms the hyoid apparatus in the adult, which supports the laryngeal structures and serves as the base of the tongue (Duellman and Trueb, 1986).

1.3.2. Reproduction

Despite the great number of specializations in life history, including terrestriality and direct development, the reproduction of anurans is always linked to water (Vitt and Caldwell, 2009). In the tropics, anurans are, in principle, capable of reproducing throughout the year, the main limiting factor being rainfall; at higher latitudes and higher altitudes, temperature increasingly becomes and important factor controlling time of breeding and length of the breeding season, and thus, the reproductive activity in temperate species is typically seasonal (Duellman and Trueb, 1986). Other factors for reproduction might include stimulus from light intensity over short time periods, as in the pipid frogs *Hymenochirus boettgeri* and *Xenopus laevis* (Rabb and Rabb, 1963; Savage, 1965).

In addition to reproductive organs, external phenotypic differences exist between males and females in most anurans. Sometimes they persist throughout the year, or they can appear only in the breeding season. This sexual dimorphism is reflected in differences in size, glandular development, skin texture, dermal ornamentation, vocal sacs, presence of tusks or spines, and coloration (Duellman and Trueb, 1986). The most notable secondary sexual characters, except for vocal sacs, are the nuptial excrescences (modified dermal and epidermal tissues) on the prepollices of males during the breeding season, whose main function is to firmly hold the female during the amplexus, even though they may also play a role in male-male combat (Duellman and Trueb, 1986). Breeding males of some species may produce adhesive substances in abdominal and other glands to facilitate the amplexus (Duellman and Trueb, 1986), and breeding males of *Hymenochirus* and *Pseudhymenochirus* display conspicuous postaxillary glands (Rabb and Rabb, 1963).

During reproduction, most frogs produce the amplexus behaviour, in which the male grasps the female so that his cloaca is positioned just above the cloaca of the female (Vitt and Caldwell, 2009). Sometimes the amplexus stimulates the ovulation but is not necessary for fertilization, which occurs after the amplexus (Duellman and Trueb, 1986). In most non-neobatrachian, and some basal neobatrachian frogs (Myobatrachidae, Telmatobiinae, and Sooglossidae) the amplexus is inguinal, while most derived neobatrachians have an axillary amplexus, even though there are notable exceptions in both cases (Duellman and Trueb, 1986). The inguinal amplexus is not as efficient as the axillary because vents are not juxtaposed in the former (Rabb, 1973; Duellman and Trueb, 1986). Superimposed to the general pattern of inguinal amplexus in non-neobatrachian and basal neobatrachian frogs and axillary amplexus in derived neobatrachian frogs, there are several variants related to the relative body size and shape of the sexes, and the model of ovoposition (among others). Other variants include (i) *Alytes obstetricans* (Alytidae), where the amplexus is inguinal and then the male moves forward to a axillary position; (ii) species with cephalic amplexus (*Colostethus inguinalis*; Dendrobatidae), (iii) species that become glued by adhesive substances produced by the dermis (*Breviceps adspersus*; Brevicipitidae); (iv) staddle amplexus (*Guibemantis*

liber; Mantellidae) (Vitt and Caldwell, 2009). In most species, fertilization and ovoposition occur while the pair is in amplexus; some species can postpone fertilization; and others display a complex mating behaviour, such as the dances performed by the members of the aquatic family Pipidae (Duellman and Trueb, 1986).

The major factor in anuran courtship behaviour is the production of advertisement calls by males, and olfactory and visual clues seem to be less important (Duellman and Trueb, 1986). However, olfactory clues have been shown to be very important in aquatic species: in *Hymenochirus boettgeri*, secretions from postaxillary glands may attract females or repel other males, and because they are distinctively coloured, they may also act as visual clues (Rabb and Rabb, 1963).

1.3.3. Sound production

Vocalization plays a central role in the mating and territorial behaviour of frogs, which use their calls to delimit territories and attract mates (Duellman and Trueb, 1986; Márquez et al., 2001). Advertisement calls are species-specific and being subjected to a strong selective pressure by female preferences, they constitute a major pre-mating reproductive isolation mechanism in anurans (Márquez et al., 2005). Therefore, it is not surprising that such a critical function has been the subject of intensive selection through evolutionary history, and that frogs have evolved complex vocal structures capable of producing a wide variety of advertising sounds. Along with vocalization mechanisms, a sophisticated acoustic reception system (in many cases they develop an external tympanum) allows frogs to discriminate among species and individuals (Vitt and Caldwell, 2009). The diversification of sound production mechanisms is intimately linked to and constrained by the evolution of vocal structures, which is necessarily connected to the evolution of the respiratory system.

Despite the considerable diversity of calls and larynx morphologies among extant frogs, the majority of the species call by moving air from the lungs through the glottis (Fig. 1.7) (Duellman and Trueb, 1986). In most frog species, the laryngeal apparatus, which is suspended between the posteromedial processes of the hyoid (= thyrohyals), is a cartilaginous capsule composed of two arytenoid cartilages (each bearing one vocal cord), the cricoid cartilage and associated musculature (Fig. 1.7) (Duellman and Trueb, 1986). The larynx is located between the lungs and the buccal cavity, and sound production occurs by the pulmonary pressure, which forces the separation of the arytenoid cartilages, leaving the air to pass over vocal cords and associated musculature, causing them to vibrate (Duellman and Trueb, 1986). Both, males and females possess functional laryngeal apparatuses, but in males, they are larger relative to body size and possess stronger musculature (Duellman and Trueb, 1986). Moreover, only males have vocal sacs. Vocal sacs primarily serve as resonating chambers but can also act as sound-couplers or acoustic radiators to the air around them (Duellman and Trueb, 1986).

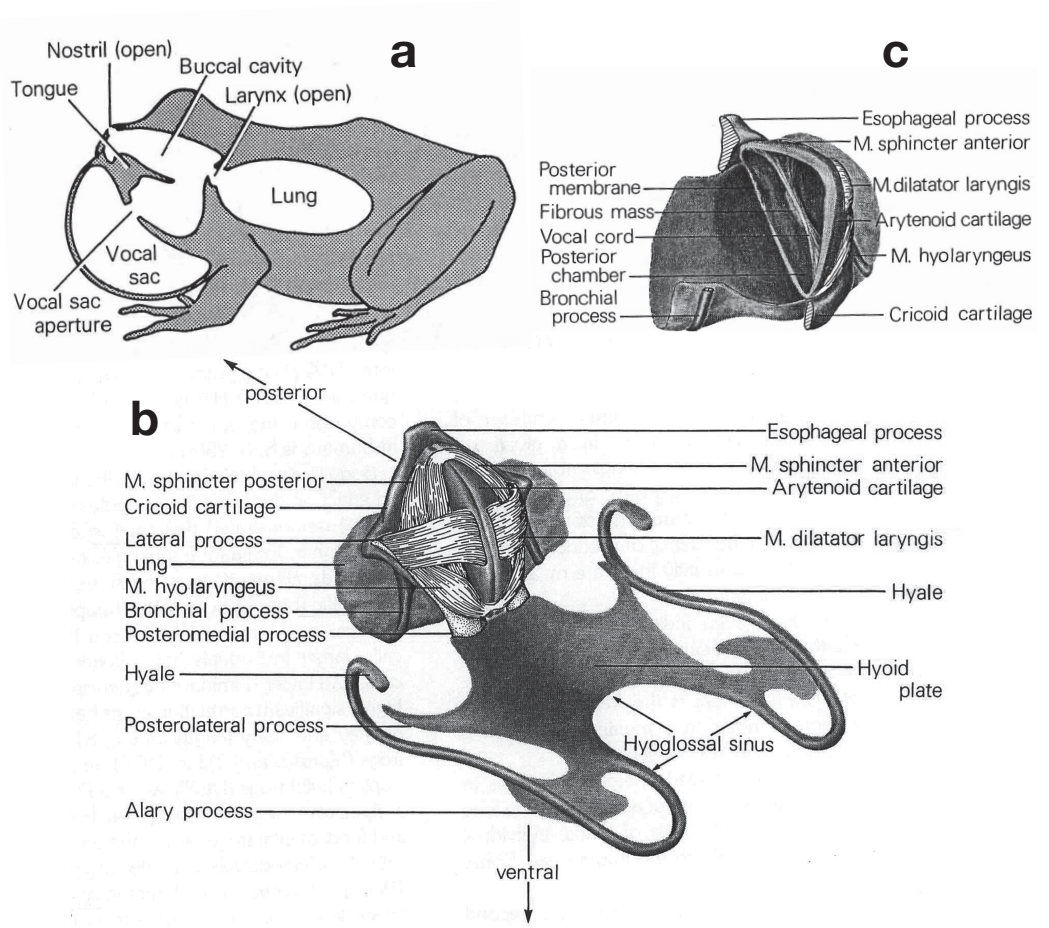


Fig. 1.7. General mechanism of sound production and associated anatomy in Anura. (a) Anatomical structures involved in vocalization. (b) Generalized representation of a hyolaryngeal apparatus. (c) Cross-sectional view of the hyolaryngeal apparatus. Modified from Duellman and Trueb (1986).

A remarkable exception to the above-described general sound production and larynx morphological patterns occurs in the family Pipidae (Yager, 1996), with the structure and function of their larynx being radically different from those of other frogs (Fig. 1.8) (Rabb, 1960; Yager, 1996). This sound production mechanism is likely another adaptation to their fully aquatic lifestyle (Cannatella and Trueb, 1988a). Pipids lack vocal cords, and their larynx is a greatly enlarged and (at least partially) ossified box made up by the cricoid cartilage and the thyrohyals, which do not form part of the larynx in non-pipid frogs. This box encloses the arytenoid cartilages, which are modified into two bony rods (Fig. 1.8) (Duellman and Trueb, 1986).

The sound production mechanism was described in detail for *Xenopus borealis* (Yager, 1992, 1996), and it appears to be based on implosion of air into a vacuum formed by rapidly moving disk-like enlargements of the arytenoids. The sound is then amplified by the enlarged voice box

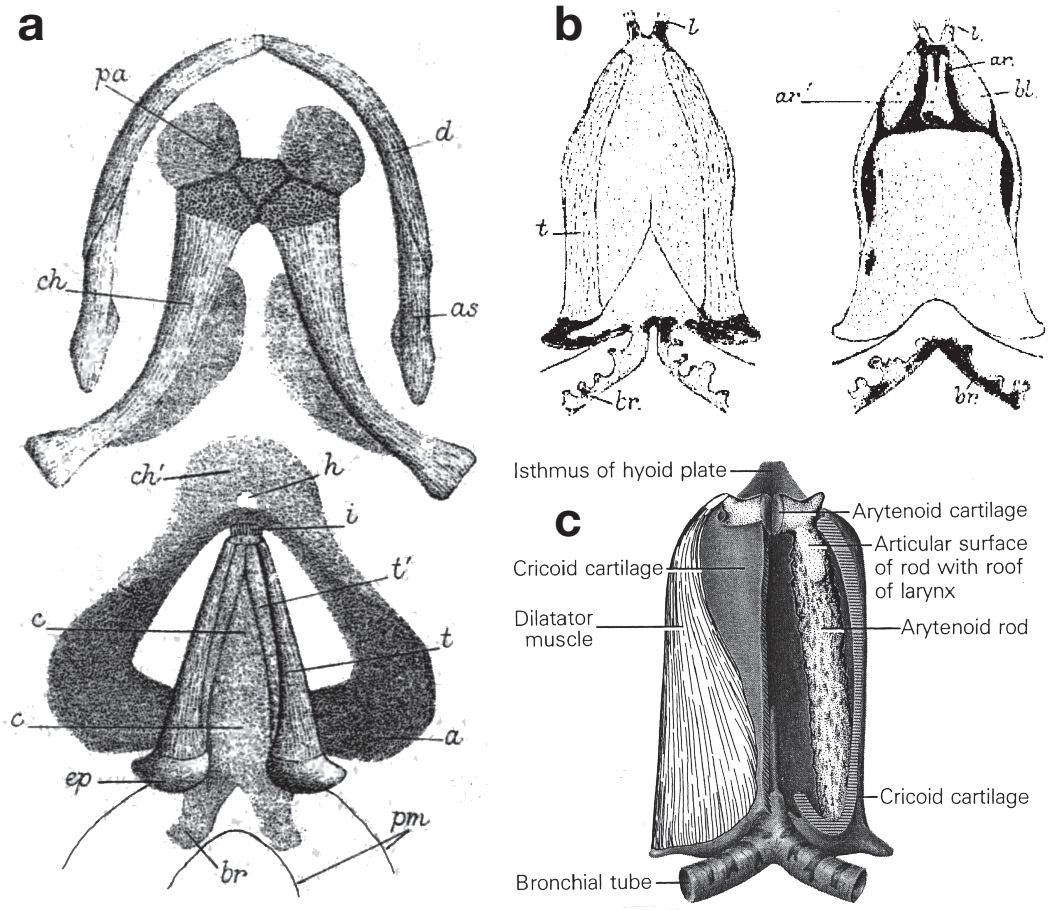


Fig. 1.8. The larynx apparatus in Pipidae. (a) *Hymenochirus boettgeri,* male. Hyobranchial, bandibular, and laryngeal skeleton in ventral view (modified from Ridewood, 1900). (b) *Xenopus laevis,* male. Laryngeal skeleton in ventral (left) and dorsal (right) views (modified from Ridewood, 1987). (c) *Pipa pipa*, male. Laryngeal apparatus in dorsal view (modified from Duellman and Trueb, 1986). Abbreviations: a, alary processes of the hyobranchial skeleton; ar, arytenoid cartilages; bl, dorsal extremity of the cricoid cartilages; br, bronchial cartilage; c, cricoid cartilage; ch, ceratohyals (= hyoidean cornu); ch', median cartilage formed by the secondary union of the ceratohyals; d, dentary bone; ep, posterior epiphyisis of the thyrohyal; h, hyoglossal foramen; i, itsmums between the anterior and posterior portion of the branchial skeleton; pa, anterior plate of the hyoidean skeleton; pm, outline of the lungs; t, thyrohyal bones; t', lamella of cartilage projecting from the thyrohyal bone in the male.

that serves as an internal vocal sac (Yager, 1992, 1996). Sound thus is produced without moving an air column, and therefore without externally visible movements of the flanks or throat. Similar motionless calling was also observed in *Hymenochirus boettgeri* (Rabb and Rabb, 1963), *Pipa pipa* (Rabb, 1960), *Pipa carvalhoi* (Weygoldt, 1976), *Xenopus laevis* (Müller and Scheer, 1970), and most other pipids (Rabb, 1969; Kunz, 2003). However, *Pseudhymenochirus* was stated to produce sounds by a more conventional sound production mechanism based on air movement (Yager, 1996), although this behaviour has not been documented in detail so far.

1.3.4. Species diversity

The highly derived anuran bauplan seems to have opened up a plethora of adaptive possibilities, as judged by the current success of anurans compared to the other two orders of living amphibians (Duellman and Trueb, 1986). Anurans have undergone a phylogenetic and ecological diversification that far surpasses that observed in salamanders and caecilians, and frogs are currently more diverse, widespread, and numerous (Duellman and Trueb, 1986). Their small body size and rapid rates or reproduction enable them to populate new areas quickly, while their low metabolic rate and capacity to adjust to wide ranges of temperature make it possible for them to adapt rapidly to environmental differences in time and space (Carroll, 2009). Despite their conserved body plan primarily adapted for jumping, anurans have evolved locomotor adaptations for both terrestrial (*e.g.,* arboreal habits, burrowing) and aquatic habitats (Handrigan and Wassersug, 2007).

Frogs and toads occur worldwide on all continents (except Antarctica) and on most continental islands, from lowlands to high elevation. They can live in a wide range of different habitats, including cold and arid environments, but are generally absent from estuarine and marine environments as they cannot tolerate salinity (Duellman and Trueb, 1986; Vitt and Caldwell, 2009).

Currently, approximately 6,000 species of anurans are recognized, in contrast to the *ca.* 620 species of salamanders and *ca.* 190 of caecilians (Amphibiaweb, 2011; Frost, 2011). The high species diversity of anurans is unevenly distributed both across the phylogeny and geographically (Fig. 1.9). Neobatrachian frogs harbour the vast majority of species diversity (> 96%), and are mainly distributed through landmasses derived from Gondwana (Duellman and Trueb, 1986). In contrast, all other non-neobatrachian frogs represent less than the 4% of anuran diversity (Table 1.1; Frost, 2011), and are primarily distributed in Laurasian landmasses (Duellman and Trueb, 1986). Within Neobatrachia, basal lineages (Heleophrynidae, Calyptocephalellidae, Myobatrachoidea and Sooglossoidea) are species-poor (2.3% of the total neobatrachian diversity) and have a relict distribution (Amphibiaweb, 2011); whereas Nobleobatrachia (3,342 species) and Ranoides (2,250 species) are highly speciose clades (Table 1.1; Frost, 2011) and their centres of diversity are located in the Neotropics and Old World, respectively (Hoegg et al., 2004; Frost, 2011).

1.3.5. Biogeography

Feller and Hedges (1998), using data from two mt rRNA and two mt transfer RNA (tRNA) genes, reconstructed a molecular phylogeny of living amphibians, where salamanders and caecilians were sister-group, to the exclusion of frogs ("Procera" hypothesis). According to these authors, this hypothesis is supported by the fossil record, as oldest fossils of both salamander and caecilians are known from the Jurassic (Jenkins and Walsh, 1993; Milner, 1993), while anurans are already

known from much earlier, the Triassic (Rage and Roček, 1989; Evans and Borsuk-Bialynicka, 1998). According to the chronology provided by the fossil record and the current patterns of distribution of amphibians (most caecilians and neobatrachian frogs are distributed in Gondwanan-derived continents, whereas salamanders and non-neobatrachian frogs in Laurasian-derived ones), Feller and Hedges (1998) suggested a biogeographic scenario where the initial breakup of Pangaea in the Mesozoic (195–157 mya; Hallam, 1994) might represent a major vicariant event between salamanders and caecilians and between reciprocally monophyletic archaeobatrachian and neobatrachian frogs. Therefore, diversification of living frogs would be post-Pangaean.

However, many recent morphological and molecular studies have pointed out to a closer relationship between salamanders and frogs, and to the paraphyly of "Archaeobatrachia" (*e.g.,* Haas, 2003; Roelants and Bossuyt, 2005; San Mauro et al., 2005; Roelants et al., 2007). Moreover, methods that are able to estimate divergence times from molecular data have provided strong evidence that splits among the three living amphibian orders, and among neobatrachian and non-neobatrachian frog lineages, predated the initial breakup of Pangaea (San Mauro et al., 2005; Roelants et al., 2007). This is further supported by other data from distribution of both fossil and living amphibians (Estes and Reig, 1973; Rage and Roček, 1989; Jenkins and Walsh, 1993; Evans et al., 1996; Evans and Borsuk-Bialynicka, 1998), including (i) the exclusive distribution of pipid frogs in Gondwanan-derived landmasses, and (ii) the disjoint distribution of most basal frogs *i.e., Ascaphus* in North America (Laurasia) and *Leiopelma* in New Zealand (Gondwana) (Green and Cannatella, 1993).

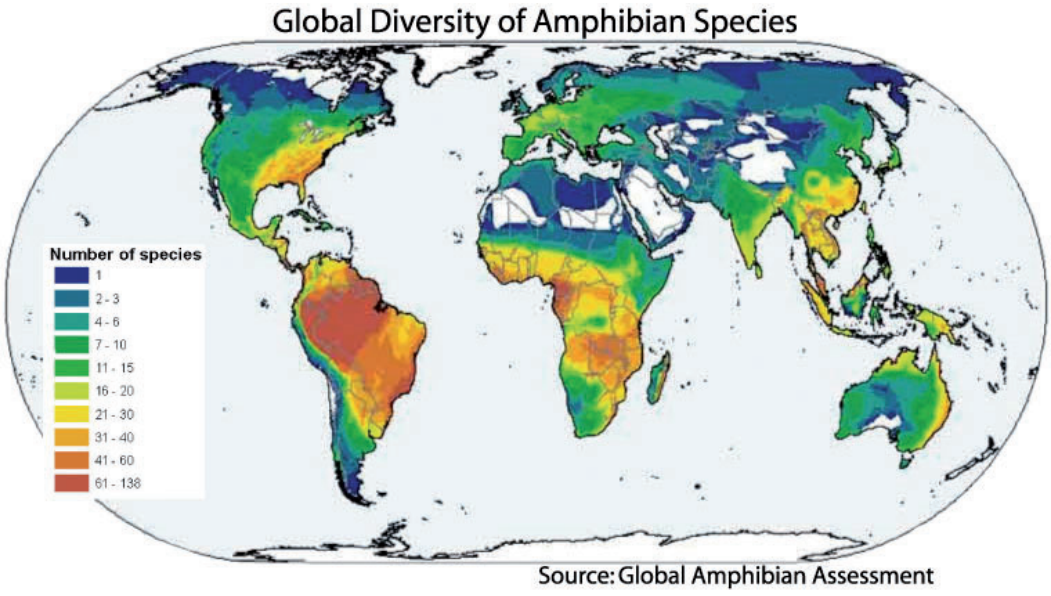


Fig. 1.9. World map showing the current diversity of amphibians (number of species). Modified from Amphibiaweb: Information on amphibian biology and conservation. [web application]. 2012. Berkeley, California: AmphibiaWeb. Available: http://amphibiaweb.org/. (Accessed: April 26, 2012).

1.3.6. Conservation

Collectively, amphibians are of global conservation concern. According to the International Union for Conservation of Nature (IUCN), approximately one-third (1,856 species; iucnredlist.org/amphibians) of all amphibian species are threatened. In the last two decades, there have been an alarming number of extinctions: nearly 168 species are believed to have gone extinct and at least 2,469 more (> 40%) have populations in decline. This greatly exceeds by far the trends observed in other groups of vertebrates (Amphibiaweb, 2011), and recent analyses suggest that the number of extinct and threatened species will probably continue to rise (Stuart et al., 2004).

The most important factor leading to amphibian population declines, as in many other endangered species, seem to be habitat destruction, alteration and fragmentation (Marsh and Trenham, 2001; Amphibiaweb, 2011), affecting nearly 4,000 amphibian species, according to the IUCN. Other threats include pollution and chemical contaminants (Blaustein and Wake, 1990; Hayes et al., 2002), introduced species (Kats and Ferrer, 2003; Vredenburg, 2004), over-exploitation (Jennings and Hayes, 1985), climate change and increased ultraviolet B radiation (Kiesecker et al., 2001; Carey and Alexander, 2003; Hof et al., 2011), and emerging infectious diseases (Daszak et al., 2003), particularly the chytrid fungus *Batrachochytridium dendrobatidis* (James et al., 2009). According to IUCN, disease appears to affect a smaller number of species, although for those species affected, it can produce dramatic population declines and rapid extinction. In contrast, habitat loss and degradation affects many more species, but the rate of decline is usually slower. There is not a single cause for amphibian declines, and the underlying mechanisms behind all the above factors are complex and may interact synergistically with each other (Kiesecker et al., 2001).

1.4. Molecular phylogenetics

Evolution is defined as descent with modification from a common ancestor (Darwin, 1859), which implies that evolutionary relationships between organisms can be represented by means of phylogenies. Nowadays, methods of phylogenetic inference are used to reconstruct evolutionary trees of genes and organisms, but their applications go far beyond pure systematics and into most branches of biology, including detection of orthology and paralogy, divergence time estimation, reconstruction of ancestral sequences, the study of selection at the molecular level, detection of recombination, among-species comparative methods, population dynamics, or evolutionary epidemiology and medicine (Felsenstein, 1985a; Huelsenbeck and Rannala, 1997; Holder and Lewis, 2003; Yang and Rannala, 2012). Phylogenetic inference is an estimation procedure of the evolutionary history, which is necessarily based on the incomplete information currently available

to us (Swofford et al., 1996). In the context of molecular phylogenetics, information is only available from contemporary species, or exceptionally, from relatively young fossils (Pääbo, 1989). Phylogenies have been long used by systematists ever since they were discussed by Darwin and Haeckel, but modern algorithmic methods did not appear until approximately 50 years ago (Felsenstein, 2004).

Besides the work by Sokal and Sneath on numerical taxonomy (*e.g.,* Sokal and Sneath, 1963), the first true phylogenetic methods aimed to be applied to molecular data were those developed by Edwards and Cavalli-Sforza, who introduced for the first time the parsimony method (1963), as well as the maximum likelihood approach (1964). This latter paper is remarkable as it supposed the first probabilistic approach to phylogenetics (Felsenstein, 2004). Nevertheless, the first reasonably complete account for the parsimony method was presented by Camin and Sokal (1965), being the reference for further work on parsimony (Felsenstein, 2004). Eck and Dayhoff (1966) first used molecular data to infer phylogenetic relationships using parsimony, shortly after first protein sequences became available. Later on, Fitch and Margoliash (1967) defined a weighted least square method and used it for the first time with molecular data, while popularizing the use of distance methods in phylogenetics. However, Fitch and Margoliash (1967) used uncorrected distances that did not account for the multiple substitutions that are likely to occur in the sequences. The first model of sequence evolution was that of Jukes and Cantor (1969), and allowed to correct multiple substitutions in distance-based estimations. Kluge and Farris (1969) and Farris (1970) introduced unordered (Wagner) parsimony that (in contrast to Camin and Sokal, 1965) did not assume irreversibility. Fitch (1971) introduced unordered parsimony for molecular data.

1.4.1. Algorithms versus optimality criteria

Phylogenetic inference methods can be divided into those that use (i) algorithms or (ii) optimality criteria. Algorithmic methods build a tree following a specific sequence of steps, combining tree inference and the definition of the preferred tree into a single statement (Swofford et al., 1996). Algorithms include distance-based methods, such as least squares (Cavalli-Sforza and Edwards, 1967), and Neighbour-joining (Saitou and Nei, 1987), as well as most first-generation parsimony methods (Kluge and Farris, 1969). Distance methods first calculate pairwise distances between all sequences to build a distance matrix, and then construct a phylogenetic tree from that matrix (Swofford et al., 1996). Algorithms have the advantage of being fast and might find trees that are close to the optimal if the data are clean (Swofford et al., 1996; Wiley and Lieberman, 2011). However, they fail to provide a sense of strength on the obtained estimate, they do not account for the high variances of large distance estimates, and more importantly, they loss lots of information when the alignment is translated into distances (Swofford et al., 1996; Yang and Rannala, 2012).

Methods based on optimality criteria, in contrast, make use of the information contained in the sequences more efficiently by considering each site in the alignments, and have the advantage of providing some measure of the certainty of the obtained results (Swofford et al., 1996; Wiley and Lieberman, 2011). Four main methods can be distinguished among those that use optimality criteria: minimum evolution (Rzhetsky and Nei, 1992), maximum parsimony, maximum likelihood and Bayesian inference (Swofford et al., 1996). They all proceed first by generating trees using algorithms, and then, evaluate those trees according to an objective function in order to choose the best tree (Swofford et al., 1996). In theory, the tree with the best score should be identified by comparing all possible trees, but because the number of possible trees is generally huge (Felsenstein, 1978b), exhaustive tree searches are generally not feasible, and heuristic tree search methods are used (Swofford et al., 1996). Heuristic approaches generate a starting tree using a fast algorithm and then perform local rearrangements to improve the tree score, and although they do not guarantee to find the best tree, they make possible to visit many trees in large datasets, increasing the chance of finding the best tree (Yang and Rannala, 2012).

1.4.2. Maximum parsimony

Maximum parsimony has been the most widely used method from the 1970s until relatively recently. The maximum parsimony criterion prefers simple explanations of data, *i.e.,* shared characters are assumed to be due to common ancestry unless they conflict which each other, in which case homoplasy must be assumed (Swofford et al., 1996). Therefore, the tree with the fewest number of independent origins of shared characters (called steps) is preferred; in other words, the most parsimonious tree is the shortest tree (Wiley and Lieberman, 2011).

Parsimony comprises a group of related methods that try to minimize tree length, but differ in their underlying evolutionary assumptions (Wiley and Lieberman, 2011). Some of them allow reversibility of change between states in an ordered (Fitch parsimony; Fitch, 1971) or unordered manner (Wagner parsimony; Kluge and Farris, 1969; Farris, 1970). Others assume some degree of irreversibility: complete irreversibility in the Camin-Sokal parsimony (Camin and Sokal, 1965), and the inability to regain a derived state once being lost in the Dollo parsimony (Farris, 1977). However, the most widespread method is the generalized parsimony (Swofford and Olsen, 1990), which allow to use all the above methods within the same analysis for different characters, which is done by means of cost matrices (Wiley and Lieberman, 2011).

The goal of minimizing evolutionary change is usually defended on philosophical grounds. According to this view, the phylogeny cannot be estimated using statistics, as it is the result of an historical non-repeatable event, and thus, one should stick to the most parsimonious explanation for the observed data (Kluge, 1979). Sometimes, maximum parsimony has been suggested to provide greater explanatory power than other methods because it minimizes the number of *ad hoc* hypotheses to explain the data (Farris, 1983). Related arguments have focused on the concept of falsability (after the works by Karl Popper), suggesting that parsimony is the only method consistent with the hypothetico-deductive framework, which should be the one preferred to test hypothesis in science (Gaffney, 1979). However, further work has shown that a careful interpretation of Popper's work does not unambiguously lead to the parsimony principle (de Queiroz and Poe, 2001; Felsenstein, 2004). In contrast, other authors have supported the maximum parsimony method for reasons other than philosophical, such as the predictiveness of the classifications that are derived from it (Beatty and Fink, 1979), or even from a statistical viewpoint (Sober, 1988). The practice of maximum parsimony has also been encouraged because it makes few assumptions about the evolutionary process, in contrast to statistical approaches that would require too much knowledge about the details of the evolutionary process (Farris, 1983).

Parsimony was long thought to make only noncontroversial assumptions of the underlying evolutionary process, but when it is examined from a statistical viewpoint it happens not to be the case, as they are implicit assumptions about rate of change in different lineages (Felsenstein, 2004). Maximum parsimony requires the reconstruction of ancestral states to find the shortest tree, but in doing so, it ignores branch lengths despite the evidence that changes are more likely to have occurred on long branches than in a short ones (Yang, 2006). Furthermore, maximum parsimony makes the unrealistic assumption of equal probability of change between states of characters (nucleotides or amino acids) (Yang, 2006). More realistic assumptions might be made by means of step matrices that decrease the weight of frequent changes (*e.g.,* transitions), but determining appropriate weights may be nontrivial and attempts to derive appropriate models naturally lead to model-based statistical methods (Yang, 2006). Another assumption of parsimony, which is shared by model-based statistical methods, is the independence of sites, which is assumed mainly for computation reasons (Wiley and Lieberman, 2011). Felsenstein (1978a) showed that maximum parsimony is an inconsistent estimator of the phylogeny under the precise conditions where parallel changes exceed informative nonparallel changes, an effect known as the long-branch attraction. These conditions are met on a particular location of the tree space, where maximum parsimony will recover the wrong tree consistently, *i.e.,* it would converge to the wrong tree as more information is added (Felsenstein, 2004). According to Felsenstein (2004), maximum parsimony is a fairly well-behaved method when it is not in a situation that involves long-branch attraction, and has the advantages of being simpler and faster than maximum likelihood.

1.4.3. Maximum likelihood

Maximum likelihood might be seen as a natural extension of the maximum parsimony method, when we want to account for differences in branch lengths and substitution rates between nucleotides or amino acids (Yang, 2006). In fact, maximum parsimony and maximum likelihood have been shown to be equivalent under a specific set of parameters (Tuffley and Steel, 1997). The application of the likelihood principle to phylogenetics (Edwards and Cavalli-Sforza, 1964; Felsenstein, 1981) allowed studying evolutionary relationships as a statistical rather than philosophical problem, allowing a framework for estimating historical patterns, inferring intrinsic parameters of evolutionary processes, and testing hypotheses under the auspices of the neutral theory of molecular evolution (Kumar et al., 2012). The likelihood principle is defined as the probability that a proposed model of evolution (for either nucleotides or amino acids) and a hypothesized evolutionary history (tree) would give rise to the observed data (sequences) (Swofford et al., 1996). Thus, maximum likelihood assumes that the model of evolution being used is correct (Swofford et al., 1996). Models are probabilistic descriptions of the evolutionary process: changes between nucleotides (or amino acids) are described by continuous-time Markov chains, where nucleotides (or amino acids) are states of that chain (Yang, 2006). The most important property of Markov models is that the change only depends on the current state (it has no memory), and they also assume that sites are evolving independently (Yang, 2006). We often place further constrains, giving rise to a variety of evolutionary models, which differ in their assumptions about the evolutionary process (Yang, 2006).

Evolutionary models are essential for the statistical inference of phylogenies, as they permit calculating the probabilities of change between states of character, but given the variety of available models, an objective criterion is needed for model selection, which happens to be the maximum likelihood (Posada, 2009). The most appropriate evolutionary model is the one with the highest likelihood score. This choice has been traditionally performed through hierarchical likelihood ratio tests that compare models of increasing complexity in a pairwise, nested manner (Posada and Buckley, 2004). However, other alternatives such as the Akaike information criterion (AIC; Akaike, 1973) or Bayesian information criterion (BIC; Schwarz, 1978) have been favoured later, as they avoid overfitting the data with too many model parameters relative to the increase of the likelihood score (Posada and Buckley, 2004). In addition, hierarchical likelihood ratio tests cannot be applied to amino acid replacement matrices.

Fisher (1922) showed that maximum likelihood has a variety of good asymptotic properties for estimating parameters from a model, including consistency (converging to the correct value of the parameter) and efficiency (having the smallest possible variance around the true parameter value), as the amount of data grows large (Felsenstein, 2004). This is the case of maximum likelihood

estimates (MLE) of parameters, that is, the values of the model parameters that maximize the likelihood function, which are generally found iteratively with optimization algorithms (Yang and Rannala, 2012). The tree score using the maximum likelihood function is first calculated for each site, as the sum of the probabilities of every possible reconstruction of ancestral states given some model of evolution (Swofford et al., 1996). Then, assuming that sites evolve independently, the likelihood of the full tree is calculated as the product of the likelihoods for each site (Swofford et al., 1996). However, the resulting likelihood values are very small, and thus are generally expressed as the natural logarithm (-*ln*L). Two steps are involved in maximum likelihood tree estimation: first, branch lengths are optimized to calculate the tree score of each candidate tree; and then, heuristics are used to explore the tree space and find the best tree that maximizes the likelihood function (Yang and Rannala, 2012). It must be noted that from a statistical viewpoint, the tree topology is a model, whereas branch lengths, substitution rates and equilibrium frequencies are parameters of the model; thus maximum likelihood is equivalent to comparing many models, each with the same number of parameters (Yang and Rannala, 2012).

An important distinction between maximum parsimony and maximum likelihood approaches is that maximum parsimony first infers ancestral states from contemporary data, and treats these as fixed and known, without error, in further analyses (Pagel, 1997). In contrast, maximum likelihood avoids reconstructing ancestral states by integrating all possible values at ancestral nodes in the form of probabilities, which will obviously be higher for more likely combinations of characters (Pagel, 1997). It also avoids the situations where a preferred tree needs to be selected among several equally parsimonious trees (*i.e.,* trees with the same length), which is common in maximum parsimony (Swofford et al., 1996; Pagel, 1997). The maximum likelihood method is consistent (Rogers, 1997; Felsenstein, 2004), and tend to be robust to many violations of the assumptions made by the models (Swofford et al., 1996). Moreover, although the asymptotic theory of MLE cannot be applied to tree reconstruction, simulations have shown that maximum likelihood has higher efficiency than maximum parsimony or distance methods in obtaining the *true* tree (Felsenstein, 2004; Yang and Rannala, 2012). Furthermore, all model assumptions are explicit in maximum likelihood, and so they can be easily evaluated and improved (Swofford et al., 1996). Currently, a wide repertory of sophisticated evolutionary models is available (Yang and Rannala, 2012). The main drawbacks of model-based methods are that they require significantly more computational power, and may behave poorly if the used model is grossly incorrect (Yang and Rannala, 2012).

One of the advantages of working in a probabilistic framework is that many statistic tools can be applied to *a posteriori* analysis of reconstructed phylogenies. In particular, tests of alternative phylogenetic hypotheses previously suggested on the literature are very useful to assess whether analyzed data, besides supporting the most likely hypothesis, could support or reject other hypotheses, specially those based on morphology (Schmidt, 2009).

1.4.4. Bayesian inference

Bayesian statistics have a fundamental difference compared to classical, or frequentist, approaches (to which maximum likelihood belongs): model parameters are considered random variables with statistical distributions rather than unknown constants, as in maximum likelihood (Yang and Rannala, 2012). Bayesian proponents argue that since the value of the parameter is unknown, it is sensible to specify a probability distribution to describe its possible values (Huelsenbeck and Rannala, 2004; Yang, 2006). The maximum likelihood approach seeks to find the highest point in the parameter space, whereas Bayesian inference measures the volume under a posterior probability surface, so as that the nuisance parameters (*i.e.,* those that need to be estimated but are not of interest) are integrated out to obtain the marginal posterior probability (Holder and Lewis, 2003). Bayesian inference is based on Bayes's theorem, which states that the posterior probability is proportional to the prior probability multiplied by the likelihood. It is expressed using the following formula:

$$P(T,\theta \mid D) = \frac{P(T,\theta) \bullet P(D \mid T,\theta)}{P(D)}$$

where $P(T,\Theta)$ is the prior probability for tree T and parameter Θ; $P(D|T,\Theta)$ is the likelihood or probability of the data given the tree and parameter; and $P(T,\Theta|D)$ is the posterior probability (Yang and Rannala, 2012). The denominator, P(D) is the marginal probability of the data, and it is a normalizing constant whose only role is to ensure that $P(T,\Theta|D)$ sums over the trees and integrates over the parameters to one (Yang and Rannala, 2012). The marginal probability of the data is a sum over all possible tree topologies; and for each topology, an integral over all branch lengths and substitution parameters (Yang, 2006).

Bayesian inference was introduced to molecular phylogenetics in the late 1990s (Rannala and Yang, 1996; Mau and Newton, 1997; Yang and Rannala, 1997; Li et al., 2000), but the early methods assumed a molecular clock (Yang and Rannala, 2012). Latter implementations eliminated the molecular clock constraint and developed more efficient Markov chain Monte Carlo (MCMC) algorithms, popularizing Bayesian inference methods (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Before the analysis, all parameters are assigned a prior distribution, which is combined with the data to obtain their posterior distribution by means of the above formula, and all inferences are based on these posterior distributions (Yang, 2006). Computationally, posterior probabilities cannot be directly calculated except for very simple cases, and thus MCMC algorithms (Metropolis et al., 1953; Hastings, 1970) are used to generate a sample from the posterior distribution. The MCMC is a simulation algorithm that avoids direct calculation of the posterior probabilities, starting from a random tree with random branch lengths and random substitution parameters, and moves from one tree (or parameter value) to another (Yang and Rannala, 2012). In the long run, the MCMC algorithm visits the trees (or parameters) in proportion to their posterior probabilities (Yang and Rannala, 2012).

Bayesian methods have a strong connection with maximum likelihood, as the preferred hypothesis is that with the higher posterior probability, a value directly proportional to the likelihood (Holder and Lewis, 2003). Furthermore, Bayesian inference is based on the likelihood function, so it shares the properties of consistency and efficiency of maximum likelihood (Holder and Lewis, 2003). However, several important differences exist between maximum likelihood and Bayesian inference. Bayesian inference simultaneously estimates the tree and associated uncertainty (measured as Bayesian posterior probabilities), and the results from Bayesian inference are easy to interpret: the posterior probability of a tree is simply the probability that the tree is correct, given the data and model (Yang and Rannala, 2012). In contrast, maximum likelihood calculates the probability that a given model would have given rise to the observed data, under a given phylogenetic tree (Swofford et al., 1996). However, the difficulty of interpreting maximum likelihood comes from the fact that no method exists that can define confidence intervals for trees (Yang and Rannala, 2012). Thus, other approaches need to be used, typically the non-parametric bootstrapping, whose interpretation is elusive (Felsenstein and Kishino, 1993; Berry and Gascuel, 1996; Susko, 2010). On the other hand, posterior probabilities have been criticized of being overestimates (Huelsenbeck and Rannala, 2004; Yang and Rannala, 2005), and it might be possible that posterior probabilities are sensitive to model violations and might be thus inflated when too simplistic models are used (Suzuki et al., 2002).

Moreover, Bayesian inference requires the specification of distributions for all parameters *a priori*, which might be positive because previous knowledge can be incorporated in the analysis. However, this is generally seen as a burden to the user, and can also have an unexpected influence in the outcome (Holder and Lewis, 2003; Brown et al., 2010; Rannala et al., 2012). Prior information is rarely available, and thus most analyses are conducted using the default priors of the program that generally assume largely uninformative (flat) prior distributions, so that most of the observed differences in the posterior distributions are attributable to the likelihood (Huelsenbeck and Rannala, 2004). However, even if the prior distributions are flat, trees obtained by Bayesian inference and maximum likelihood may differ simply because the distinction between marginal and joint estimation procedures, respectively (Holder and Lewis, 2003).

Another difference is that Bayesian inference allows easier implementation of partitions in the data set, each with independent models due to the use of MCMC, which makes them computationally less demanding (Holder and Lewis, 2003). Bayesian methods, unlike maximum likelihood, require the assessment of convergence between MCMC chains to ensure the tree space has been thoroughly explored, but this task might be difficult (Holder and Lewis, 2003). However, several tools have been developed to allow checking convergence in an appropriate manner (Nylander et al., 2008; Rambaut and Drummond, 2009; Ronquist et al., 2012).

1.5. The mitochondrial genome

1.5.1. Architecture of the mitochondrial genome

The mt genome of animals is a double-stranded, covalently closed circular molecule of typically 13-20 Kb in size that encodes a set of 13 protein-coding genes, 22 tRNA genes and two rRNA genes, with few exceptions (Fig. 1.10) (Boore, 1999; Lynch, 2007). Both rRNA and tRNA genes are involved in the translation of mt proteins, whereas protein-coding genes encode subunits of the respiratory chain for oxidative phosphorylation (Boore, 1999). They are involved in complex I (*nad1, nad2, nad3, nad4, nad4L, nad5, nad6*), complex III (*cob*), complex IV (*cox1, cox2, cox3*) and complex V (*atp6, atp8*) of the respiratory chain (da Fonseca et al., 2008). These protein complexes are complemented by other subunits encoded by the nuclear genome, which is in charge of regulating the overall process as well (Scarpulla, 2008). The interaction between mt and nuclear genomes is not restricted to oxidative phosphorylation, as it has been estimated that more than 1,500 nuclear genes are involved in regulating the mt function, including DNA replication, gene expression and modulation, complex assembly etc. (Wallace, 2005). One of the most widely accepted hypothesis for the maintenance of the mt genome is to allow a direct control of the gene expression by the oxidative phosphorylation (Allen, 2003). Alternatively, it has been suggested that differences in the genetic code might prevent mt genes to be completely transferred to the nucleus (de Grey, 2005), or that the mt proteins of the respiratory chain are extremely hydrophobic to be synthesized in the cytoplasm, and thus require a different subcellular compartment (von Heijne, 1986).

In animals, the mt genome has a compact organization with very few and short non-coding regions (5-10%; Fig. 1.10) (Wolstenholme, 1992; Lynch, 2007), which is produced by the selective pressure towards size reduction (Schneider and Ebert, 2004). Protein-coding genes consist in single exons and lack introns (with the exceptions found in two cnidarians and a placozoan; Lynch, 2007). Size reduction has produced the overlap of few nucleotides between certain open reading frames (typically *atp6–atp8, nad4–nad4L* and *nad5–nad6* in the mt genome of vertebrates), and both rRNA and tRNA genes are shorter than their prokaryotic homologues, although they retain most conserved regions and secondary structure (Wolstenholme, 1992). Protein-coding genes may possess incomplete stop codons (T or TA), which presumably become functional by subsequent polyadenylation of the transcribed messenger RNA molecules (Ojala et al., 1981).

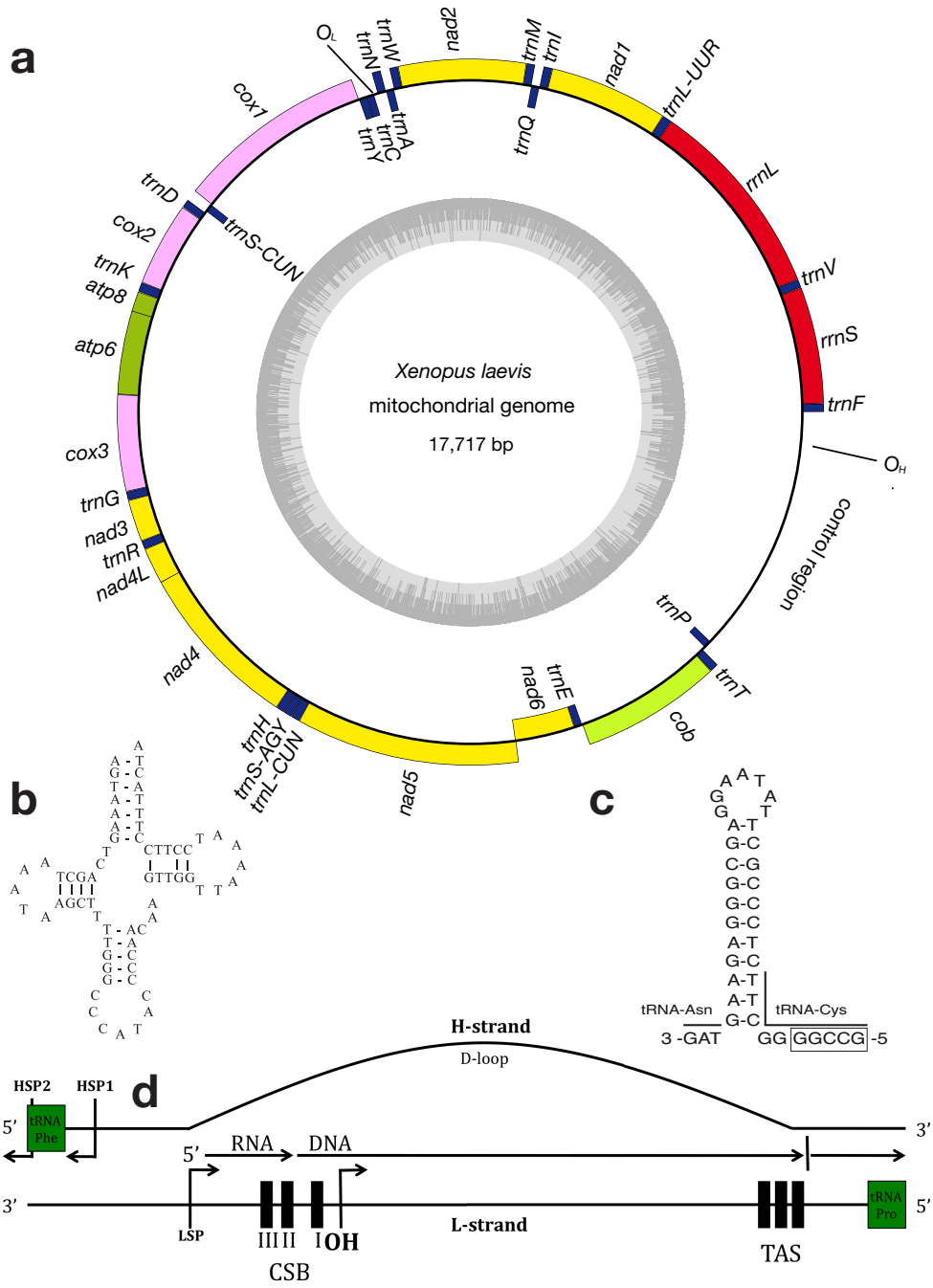


Fig. 1.10. The mitochondrial genome. (a) The mitochondrial genome of *Xenopus laevis*, as an example of the consensus gene order of vertebrates. The gene content is also the typical of metazoans, including protein-coding genes of the complex I (yellow), complex III (light green), complex IV (pink), and complex V (dark green); ribosomal RNA genes (red), and transfer RNA genes (blue); and major non-coding regions, including the origins of replication of the light (O_L) and heavy (O_H) strands. (b) Typical secondary structure of a transfer RNA gene. (c) Secondary structure of the origin of replication of the light strand. (d) Schematic representation of the control region, showing the origin of replication of the heavy strand (O_H), the conserved sequence blocks (CSB) and termination-associated sequences (TAS). Note the three-stranded structure generated by the nascent chain during replicaton. Section (d) is modified from Fonseca (2011).

In the mt genome of vertebrates, tRNA genes are interspersed between protein-coding genes (with the exception of those that overlap) (Fig. 1.10), an aspect of the mt genome architecture that could be related to the mode of transcription (Mabuchi et al., 2004). During the transcription process, a polycistronic primary transcript is generated in each strand. The two primary transcripts are later excised into mono or bicistronic mature transcripts (Wolstenholme, 1992). It has been proposed that sequences of tRNA genes provide the punctuation marks for proper cleavage of primary transcripts (Ojala et al., 1980), a mechanism supported by identifications of key enzymatic activities (*i.e.,* precise cleavage at the 5' and 3' ends of tRNA genes) in the human mt genome (Rossmanith et al., 1995). Animal mt tRNA genes show a high structural diversity, mostly due to size and sequence variations in DHU and TΨC arms, even though most tRNA genes retain the four-armed cloverleaf secondary structure, with the known exception of *trnS–(AGY)* in vertebrates (Wolstenholme, 1992; Boore, 1999). Mitochondrial tRNA genes are able to read all codons from a four-codon family, as the 22 tRNA genes are not enough to decode the 60 different codons present in the vertebrate mt genetic code (Wolstenholme, 1992).

The vertebrate mt genetic code differs from the universal code in the reassignment of three codons: AUA (isoleucine in the universal code) is reassigned to code for methionine, UGA (stop codon) is given a new sense and encodes for tryptophan, and AGA (arginine) becomes a stop codon (Wolstenholme, 1992). Besides vertebrates, at least 12 (and probably more; Abascal et al., 2006) genetic code variants occurred in bilaterian animals, with three to five differences in the meaning of codons compared to the universal genetic code (Lynch, 2007). Genetic code variants in mt genomes arose through codon reassignments (Osawa et al., 1992; Jukes and Osawa, 1993; Schultz and Yarus, 1994). These events are facilitated in mt genomes due to their specific characteristics of having small sizes, high degree of genetic linkage and high mutation rates (Lynch, 2007).

In vertebrates, the mt genome has two main non-coding regions, which are essential during replication and translation (Fig. 1.10). A long non-coding region of typically ~1 Kb in length can be found normally between the *trnP* and *trnF* genes, and it is known as the control region because it contains the signals necessary for the initiation of heavy strand replication, and transcription of both heavy and light strands (Montoya et al., 1982; Montoya et al., 1983; Clayton, 1984). The control region is very variable, but it contains (i) conserved sequence blocks (CSB; Walberg and Clayton, 1981) that participate in the formation of a proper RNA primer in the process of mt genome replication (Fernández-Silva et al., 2003), and (ii) termination-associated sequences (TAS) that regulate the replication process (Shadel and Clayton, 1997) as they have the capacity to arrest the replication of most newly initiated chains shortly after the origin of replication of the heavy strand (Doda et al., 1981; MacKay et al., 1986). The newly synthesized nascent chains remain associated to the light strand, therefore displacing the original heavy strand and creating a three-stranded structure (known as the D-loop); a structure that is repeatedly synthesized and degraded

(Clayton, 1982). The occurrences of both CSB and TAS vary among different vertebrates (*e.g.,* Zardoya et al., 1995a). A second non-coding region occurs typically between the *trnN* and *trnC* genes within the tRNA gene cluster known as the *WANCY* region (which includes the genes *trnW, trnA, trnN, trnC,* and *trnY*). This non-coding region is ~35 bp (base pairs) in length and corresponds to the origin of replication of the light strand (Martens and Clayton, 1979; Kang et al., 1997). It has the potential to form a stem-loop secondary structure when it is displaced from the heavy strand during replication (Kang et al., 1997), and the motif 5'-GCCGG-3' is involved the transition from RNA to DNA synthesis in the replication of human mt DNA (Hixson et al., 1986).

1.5.2. Molecular evolution

The mt genome of animals is almost strictly transmitted though maternal inheritance, although some cases of paternal leakage (Gyllensten et al., 1991; Kvist et al., 2003), and doubly uniparental inheritance in bivalve molluscs have also been reported (Zouros et al., 1994; Doucet-Beaupré et al., 2010). The main evolutionary consequence of exclusively maternal inheritance is that for every two copies of any nuclear gene, only one copy of any mt gene generally exists. Commonly, it is considered that the effective population size of the mt genome is one-fourth of that of the nuclear genome (*e.g.,* Moore, 1995), but this assertion is inaccurate, because it ignores (i) the selective interference due to genetic linkage in both mt and nuclear genomes, and (ii) the differential reproductive success of males and females (Lynch, 2007). However, the relative effective population sizes for mt and nuclear genomes can only be calculated through empirical observation (for example, using neutral polymorphism and divergence data), and empirical data demonstrate that the effective population size for the mt genome are smaller than their nuclear counterparts (Lynch, 2007). Recombination in animal mt genomes is very limited (Brown, 1983), and although homologous recombination activity can be present (Thyagarajan et al., 1996), the participation of both maternal and paternal genomes are required to produce genetically effective recombination (Tsaousis et al., 2005; Lynch, 2007). Hence, the role of recombination in animal mt genome is considered restricted to specific organisms, although the extent of this phenomenon is still unclear (Castellana et al., 2011).

Data accumulated for more than 30 years has shown that mt DNA evolves at a rate approximately 5 to 10 times faster than single-copy nuclear genes, although this varies extremely across genes and taxa (Brown et al., 1979; Moritz and Brown, 1987; Meyer, 1993; Lynch, 2007). The faster evolutionary rates of mt genomes are probably produced by their higher mutation rates (Denver et al., 2000; Denver et al., 2004; Lynch, 2007), although the fixation of mutations (*i.e.,* substitution events), is a complex process in which both evolutionary (selection) and demographic (drift) forces are at play (Lynch, 2007; Bromham, 2009b). During oxidative phosphorylation, mitochondria generate high levels of free oxygen radicals, producing an internal environment prone to DNA damage, including deamination of cytosine to uracil and the oxidative modification of guanine to 8-oxoguanine, which cause C:G→T:A transitions and C:G→A:T transversions, respectively (Richter

et al., 1988; Lynch, 2007). This is consistent with the high content of adenine and thymine that exists in almost all mt genomes (Perna and Kocher, 1995). Other factors that contribute to the high mutational pressure in mt DNA are the absence of histone-like proteins (Castellana et al., 2011), the inaccuracy of the DNA repair system (Bogenhagen, 1999), and the particular replication model with single-strand intermediates (Reyes et al., 1998). Additionally, mt DNA is frequently replicated within non-dividing cells (unlike nuclear DNA), thus increasing the rate of error per cell cycle (Lynch, 2007). The mt genome appears in multiple copies per mitochondrion, increasing the likelihood of new mutations to arise. It has been suggested that new mutations could be involved in gene conversion and alter the mutational profile, although almost nothing is known about the magnitude or direction of gene conversion in mitochondria (Lynch, 2007).

Because of the high mutation rates and reduced ability to shed mutations by recombination, mt genomes are expected to be susceptible to degradation by a process known as Muller's ratchet (Muller, 1964; Felsenstein, 1974). This accelerated mutational meltdown is compensated by purifying selection, which acts to conserve gene function of mt genes (Castellana et al., 2011). Cases of positive selection have also been reported in mt DNA (Meiklejohn et al., 2007; Castoe et al., 2008; da Fonseca et al., 2008).

1.5.3. Mitochondrial gene order and mechanisms of gene rearrangement

The gene order is fairly conserved among vertebrate mt genomes (Boore, 1999; Gissi et al., 2008). It has been proposed that the relative location and orientation of genes in the mt genome of vertebrates may be at an optimum state to ensure a controlled gene expression, hence favouring the stability of this particular genome architecture (Amer and Kumazawa, 2007). Nevertheless, several cases of gene rearrangements that depart from the consensus order of vertebrates have been reported in several groups, including fishes (*e.g.,* Inoue et al., 2001; Mabuchi et al., 2004), reptiles (*e.g,.* Amer and Kumazawa, 2007; Jiang et al., 2007), birds (*e.g.,* Mindell et al., 1998; Verkuil et al., 2010), mammals (*e.g.,* Janke et al., 1994), and amphibians (*e.g.,* Sumida et al., 2001; Mueller and Boore, 2005; San Mauro et al., 2006; Kurabayashi et al., 2008).

In vertebrate mitochondria, it is generally considered that gene rearrangements are the product of errors during the replication process, such as slipped-strand mispairing or asynchrony in the points of initiation and termination (Levinson and Gutman, 1987; Moritz and Brown, 1987; Mueller and Boore, 2005). Replication errors could produce that a mitogenomic region is duplicated in tandem, and redundant gene copies are expected to be lost subsequently due to selective pressure towards size reduction. Depending on which gene copies are lost, this process may produce a gene rearrangement in the duplicated block of genes. This mechanism is known as the tandem duplication–random loss model (Moritz and Brown, 1987; Moritz et al., 1987; Boore, 2000). This model has been shown to be consistent with most gene rearrangements found in vertebrates

(Moritz and Brown, 1987; Moritz et al., 1987; Macey et al., 1997; Mindell et al., 1998; Boore, 1999; Mueller and Boore, 2005; Alam et al., 2010). It is further supported by observations of non-coding regions that contain duplications in tandem, and duplicated genes or their remnants (pseudogenes) that persist after previous events of tandem duplication (Kumazawa et al., 1995; Gach and Brown, 1997; Macey et al., 1998; Mabuchi et al., 2004; San Mauro et al., 2006). Nevertheless, the tandem duplication–random loss model fails to account for gene rearrangements that involve changes in coding strands (Amer and Kumazawa, 2007) and challenges the interpretation of some gene orders with non-tandem repeats (Kurabayashi et al., 2008; Kurabayashi et al., 2010). Therefore, alternative mechanisms such as inversion through gene remoulding or intramolecular recombination (Amer and Kumazawa, 2007), or gene transposition (Macey et al., 1997), have been invoked. Specifically, intramolecular recombination can be generated either by illegitimate recombination via minicircle (Dowton and Campbell, 2001; Mueller and Boore, 2005), or by homologous recombination (Thyagarajan et al., 1996; Tsaousis et al., 2005). In any case, examples of gene inversions and non-tandem repeats are few among vertebrates, compared to the vast majority of gene rearrangements that can be successfully explained by the tandem duplication–random loss model, which is currently considered the most common mechanism for gene order change.

Earlier studies have suggested that gene duplications are more likely to occur at specific regions of the mt genome due to their particular mechanistic constraints during replication (Moritz and Brown, 1986; Mindell et al., 1998; Dowton and Austin, 1999; Boore, 2000). The random loss of duplicated genes will, in most cases, produce a new detectable gene order, and thus, gene rearrangements in these regions will appear more frequently than in others. Some previous studies have found compelling evidence that the origins of replication of both light and heavy strands are hot spots for gene order change (Mindell et al., 1998; San Mauro et al., 2006)

1.5.4. The mitogenomic approach to phylogenetics

Mitochondrial DNA has been widely used as a molecular marker during past decades, from the study of population structure, gene flow, hybridization, or biogeography, to phylogenetic relationships among species (Moritz and Brown, 1987). The conserved gene content, higher mutation rates (compared to single-copy nuclear DNA), limited rates of recombination, smaller effective population sizes (making fixations of mutations more likely in rapid speciation processes), direct orthology of genes, and the high number of copies per cell promoted the use of mt sequences over other alternatives (Curole and Kocher, 1999). The use of mt markers was also facilitated by the availability of primers to amplify by polymerase chain reaction (PCR) several regions of the mt DNA on a broad spectrum of animals (Kocher et al., 1989; Palumbi et al., 1991), although most early studies used partial fragments (~300-600 bp) of single genes (*e.g.,* Shields and Kocher, 1991; Hedges and Maxson, 1993; Montgelard et al., 1997).

Several studies that assessed the relative phylogenetic performance of individual mt genes reached contrasting conclusions, supporting that the phylogenetic potential of genes is context-specific (Russo et al., 1996; Zardoya and Meyer, 1996; Cummings and Meyer, 2005; Mueller, 2006; San Mauro et al., 2009). However, they all agreed that single mt genes might not contain sufficient information to adequately represent neither the entire mt genome nor the evolutionary history of the organisms, and could produce misleading and/ or weakly supported results. Hence, the use of complete mt genomes represents a good alternative, as it provides a large number of molecular characters to estimate statistically robust phylogenetic hypotheses.

Mitogenomics have been proven useful to reconstruct deep-level phylogenies of several animal groups over a broad timescale, up to the Permian or Carboniferous periods (~ 260-360 mya) (Zardoya and Meyer, 2001; Zhang et al., 2005a; Cameron et al., 2007; Fenn et al., 2008). Specifically, the use of complete mt genomes to reconstruct statistically robust phylogenetic relationships among (Zardoya and Meyer, 2001; Zhang et al., 2005a) and within (San Mauro et al., 2004a; Igawa et al., 2008; Zhang and Wake, 2009a, b) amphibian orders provided very successful results.

1.5.5. Mitogenomics in Anura

Twenty-three complete or nearly complete mt genomes were available for anurans at the beginning of this work in 2009 (Table 1.2). Although representatives for all five major lineages of frogs were available, the taxonomic coverage was very uneven and new mt genome data were necessary on representatives of the relict genus *Leiopelma* (Amphicoela), families Pipidae and Rhinophrynidae, and basal families within Neobatrachia, as they are crucial to address unresolved phylogenetic questions in anuran phylogeny.

Anuran mt genomes possess the archetypical gene content of most metazoans (Boore, 1999): 13 protein-coding genes, 22 tRNA genes and two rRNA genes, with very few exceptions. In *Polypedates megacephalus* (Rhacophoridae) the genes *atp8* and *nad5* are missing (Zhang et al., 2005b). The genes *trnA, trnN*, and *trnC* are absent in *Limnonectes bannaensis* (Dicroglossidae) (Zhang et al., 2009), while most mt genomes sequenced from members of the family Dicroglossidae possess an extra tRNA gene for methionine (Liu et al., 2005; Ren et al., 2009; Zhang et al., 2009; Zhou et al., 2009). The mt gene order in all non-neobatrachian frogs sequenced so far, conform to the vertebrate consensus (Fig. 1.10) (*e.g.,* Roe et al., 2005; San Mauro et al., 2004a; Gissi et al., 2006). In contrast, neobatrachian frogs exhibit a diversity of gene orders. Many neobatrachian species share a common mt gene arrangement, which is considered the consensus order of neobatrachians (Sumida et al., 2001). It departs from the vertebrate consensus mt gene order in the translocation of the *trnL–(CUN), trnT* and *trnP* genes to form the *LTPF* tRNA cluster between the control region and the *rrnS* gene (Sumida et al., 2001). Further gene rearrangements have been

found in the families Dicroglossidae (Liu et al., 2005; Ren et al., 2009; Zhang et al., 2009; Zhou et al., 2009), Ranidae (Su et al., 2007; Kurabayashi et al., 2010), Rhacophoridae (Sano et al., 2004, 2005; Zhang et al., 2005b) and Mantellidae (Kurabayashi et al., 2006). Notably, a high diversity of gene orders has been found in ranid frogs (Kurabayashi et al., 2010), and in all 12 genera of mantellids (Kurabayashi et al., 2008), including translocated genes, duplicated genes and control regions, and the presence of pseudogenes.

Table 1.2. Mitochondrial genomes of anurans available as of 2009, showing their gene content and order, as well as their RefSeq or GenBank accession number. The archetypical gene content refers to the typical 37 genes of most metazoans. Note that variations from the consensus of vertebrates both in gene content and order is by far more frequent among neobatrachian frogs.

Family	Genus	Species	RefSeq/GenBank	Gene content	Gene order
AMPHICOELA					
Leiopelmatidae	*Ascaphus*	*truei*	AJ871087	archetypical	vertebrate consensus
DISCOGLOSSOIDEA					
Alytidae	*Alytes*	*obstetricans*	NC_006688	archetypical	vertebrate consensus
	Discoglossus	*galganoi*	NC_006690	archetypical	vertebrate consensus
Bombinatoridae	*Bombina*	*bombina*	NC_006402	archetypical	vertebrate consensus
	Bombina	*maxima*	NC_011049	archetypical	vertebrate consensus
	Bombina	*orientalis*	NC_006689	archetypical	vertebrate consensus
	Bombina	*variegata*	NC_009258	archetypical	vertebrate consensus
PIPOIDEA					
Pipidae	*Xenopus*	*laevis*	NC_001573	archetypical	vertebrate consensus
	Silurana	*tropicalis*	NC_006839	archetypical	vertebrate consensus
PELOBATOIDEA					
Pelobatidae	*Pelobates*	*cultripes*	NC_008144	archetypical	vertebrate consensus
NEOBATRACHIA					
Hylidae	*Hyla*	*chinensis*	NC_006403	archetypical	neobatrachian consensus
	Hyla	*japonica*	NC_010232	archetypical	neobatrachian consensus
Bufonidae	*Bufo*	*gargarizans*	NC_008410	archetypical	neobatrachian consensus
	Bufo	*japonicus*	NC_009886	archetypical	neobatrachian consensus
	Duttaphrynus	*melanostictus*	NC_005794	archetypical	neobatrachian consensus
Microhylidae	*Kaloula*	*pulchra*	NC_006405	archetypical	neobatrachian consensus
	Microhyla	*heymonsi*	NC_006406	archetypical	neobatrachian consensus
	Microhyla	*okivavensis*	NC_010233	archetypical	neobatrachian consensus
	Microhyla	*ornata*	NC_009422	archetypical	neobatrachian consensus
Dicroglossidae	*Fejervarya*	*limnocharis*	NC_005055	2x *trnM*	derived
	Fejervarya	*cancrivora*	NC_012647	2x *trnM*	derived
	Limnonectes	*fujianensis*	NC_007440	archetypical	derived
	Limnonectes	*bannaensis*	NC_012837	2x *trnM*; lack *trnA, trnN, trnC*	derived
	Quasipaa	*spinosa*	NC_013270	2x *trnM*	neobatrachian consensus
Ranidae	*Pelophylax*	*nigromaculatus*	NC_002805	archetypical	neobatrachian consensus
	Pelophylax	*plancyi*	NC_009264	archetypical	neobatrachian consensus
	Odorrana	*tormota*	NC_009423	archetypical	derived
Rhacophoridae	*Polypedates*	*megacephalus*	NC_006408	lack atp8, nad5	derived
	Buergeria	*buergeri*	NC_008975	archetypical	derived
	Rhacophorus	*schlegelii*	NC_007178	archetypical	derived
Mantellidae	*Mantella*	*madagascariensis*	NC_007888	archetypical	derived

Besides the variations in mt gene order and content in neobatrachian frogs, they exhibit higher mt substitution rates compared to their non-neobatrachian relatives (Hoegg et al., 2004; San Mauro et al., 2004a). Yet, it is not clear when the shifts in substitution rates precisely occurred. This heterogeneous distribution of mt substitution rates among lineages of frogs, together with the high genetic divergence between frogs and their closest living sister taxa (*i.e.,* salamanders; Zardoya and Meyer, 2001) are the source of several phylogenetic artefacts in previous studies, such as the monophyly of non-neobatrachian frogs ("Archaeobatrachia": Hedges and Maxson, 1993; Hay et al., 1995) or the incorrect phylogenetic placement of Neobatrachia due to long branch attraction effects (Gissi et al., 2006). The unequal distribution of mt substitution rates across the anuran tree has also been suggested to yield considerably older time estimates for divergences among neobatrachians (Igawa et al., 2008).

1.6. Nuclear genes

1.6.1. Organization and general features

The nuclear genomes of animals are at least 100 Mb (megabase pairs) in size, and contain more than 13,000 genes, although this is very variable among species (Lynch, 2007). Nuclear DNA is organized into linear chromosomes, and most of it corresponds to non-coding DNA (90-98% in vertebrates) (Lynch, 2007). Introns are non-coding regions interspersed between exons and allow differently processing of primary transcripts through splicing (Lynch, 2007). Other important non-coding sequences are regulatory cis and trans regions upstream and downstream of genes, which are responsible of controlling transcription and processing of primary transcripts (Nei and Kumar, 2000). Many genes are part of a gene family, a group of genes (paralogs) that originated through duplication (Page and Holmes, 1998). The numerous genes present in the nucleus display various degrees of sequence conservation and a wide variety of substitution rates (slower and faster than mt genes) because they are subjected to different selective forces due to the diverse functions that the corresponding proteins are involved in (Hillis, 1987). Most animals posses two alleles per locus in the nuclear genome (*i.e.,* they are diploid), in contrast to the mt genome, which is haploid (Alberts et al., 2002), and thus recombination between homologous loci is frequent, especially during meiosis (Lynch, 2007).

In polyploid organisms, more than two alleles per loci can be found, as they possess one or more additional sets of chromosomes. These additional sets of chromosomes can originate from the same or closely related individual (autoploidization) or from the hybridization of different species

(allopolyploidization) (Otto, 2007). Allopolyploidization is a common mechanism of speciation in African clawed frogs, and several tetraploid, octoploid and dodecaploid species of *Xenopus* can be found in the wild (Evans et al., 2004; Evans, 2008). *Xenopus laevis* is tetraploid, and in addition, it is known to hybridize with three other sympatric species: with *X. gilli* in the Cape Province, South Africa, with *X. muelleri* in northern South Africa, and with *X. borealis* in Kenya (Evans et al., 2004). Polyploid species represent a clear problem to phylogenetic inference, as bifurcating trees are unable to represent the complex evolutionary history of paralogous alleles (Huber et al., 2006). However, in the present study, *Xenopus laevis* is the only polyploid species used, and because genome duplication occurred after the divergence from its sister genus *Silurana* (Evans et al., 2004), paralogous alleles present in *X. laevis* should not affect phylogenetic reconstruction.

1.6.2. Advantages and disadvantages over mitochondrial genes

The main advantage of using nuclear genes is the possibility of sampling multiple unlinked loci, in contrast to mt genes, thus increasing the potential of the data to reconstruct more robust phylogenetic hypotheses (Page and Holmes, 1998). Furthermore, the availability of nuclear genes with a wide variety of substitution rates allows their use in phylogenetics at various levels of divergence, because a marker is phylogenetically most informative when evolutionary rates are appropriate for the levels of divergence of interest (Graybeal, 1994). Specifically, nuclear genes with slower substitution rates might provide particularly useful in inferring deep-level evolutionary relationships (Springer et al., 2001).

One of the main drawbacks of using nuclear markers in phylogenetic reconstruction is the presence of recombination (Schierup and Hein, 2000). Recombination implies that different parts of a gene possess different evolutionary histories, thus sequences are not related by a single tree but rather by a set of correlated trees over the sequence (Hudson, 1983). Unlike in the mt genome (which is haploid), genealogical histories of individual nuclear loci may appear misleading or uninformative about true evolutionary relationships among organisms due to the retention and stochastic sorting of ancestral polymorphisms, a phenomenon known as incomplete lineage sorting (Maddison and Knowles, 2006). Such an effect is specially likely if the effective population sizes are large relative to divergence between the organisms at hand (Maddison and Knowles, 2006). One of the challenges of using nuclear loci in phylogenetics is the distinction of orthologous (*i.e.,* homologous through speciation) from paralogous genes (*i.e.,* homologous through duplication) within the same gene family (Gabaldón, 2008), as phylogenetic inference should only be based on orthologous characters (Fitch, 1970).

1.6.3. Nuclear genes in anuran phylogenetics

Of the nuclear markers that have been used to infer phylogenetic relationships among and within amphibian orders, the most widespread are recombination-activating gene 1 (*rag1*; Chiari et al., 2004; Hoegg et al., 2004; San Mauro et al., 2004b; Evans et al., 2005a; Roelants and Bossuyt, 2005; San Mauro et al., 2005; van der Meijden et al., 2005; Roelants et al., 2007; van der Meijden et al., 2007b; Li et al., 2009; San Mauro, 2010), recombination-activating gene 2 (*rag2*; Chiari et al., 2004; Hoegg et al., 2004; van der Meijden et al., 2005; van der Meijden et al., 2007b), chemokine receptor type 4, (*cxcr4*; Roelants and Bossuyt, 2005; Roelants et al., 2007; San Mauro, 2010), sodium-calcium exchanger 1 (*slc8a1*; Roelants and Bossuyt, 2005; Roelants et al., 2007; San Mauro, 2010), sodium-calcium exchanger 3 (*slc8a3*; Roelants et al., 2007; San Mauro, 2010), brain-derived neurotrophic factor (*bdnf*; van der Meijden et al., 2007b; Li et al., 2009), proopiomelanocortin (*pomc*; Wiens et al., 2005; Li et al., 2009), rhodopsin (*rho*; Hoegg et al., 2004; Frost et al., 2006) and histone 3 (*h3a*; Frost et al., 2006).

Recombination-activating genes 1 and 2 (*rag1* and *rag2*, respectively) form a heterodimer that is required for V(D)J recombination (V, D and J being, respectively, the variable, diversity and joining segments of the genes encoding the variable portion of the T-cell antigen receptor), which is part of the adaptive immune response in vertebrates (Schatz et al., 1989; Agrawal et al., 1998). The core region of *rag1* was derived from the *Transib* transposon superfamily, whereas the *rag2* and the N-terminal domain of *rag1* probably was derived from other sources (Kapitonov and Jurka, 2005). Both *rag1* and *rag2* genes are single-copy; they each posses single exons and are tightly linked in the genome, as evidenced by the complete genome sequence of the diploid clawed frog *Silurana tropicalis*, in which *rag1* and *rag2* exons are separated by a 6.5Kb intergenic region (Evans et al., 2005a).

The brain-derived neurotrophic factor (*bdnf*) is a member of the family of nerve growth factors, a class of molecules playing key roles in neuronal development, survival and regeneration in both the central and peripheral nervous systems (Yovanovich et al., 2009; Tettamanti et al., 2010). Members of the family of nerve growth factors evolved by two duplications at an early stage of vertebrate evolution (Tettamanti et al., 2010). The high sequence identity of the *bdnf* genes across vertebrates indicate that this gene reached an optimally functioning structure very early in evolution due to an increased selective pressure on the coding region, as compared, for example, to other members of the family such as the *ngf* (nerve growth factor) gene (Hallböök et al., 1991; Götz et al., 1992). The *bdnf* gene has a single exon of between 741 and 813 bp in length in vertebrates.

The chemokine receptor type 4 (*cxcr4*) belongs to the family of CXC chemokines, which comprises 16 ligands and six receptors in mammals (Huising et al., 2003). Among all CXC chemokine receptors, *cxcr4* is the most conserved and it is present in all vertebrates (Huising et al., 2003). The ancestral function of CXC chemokines and their receptors was probably related the development of the central nervous system, even though at present, their main role is within the immune system (Huising et al., 2003). The receptor encoded by the *cxcr4* gene is activated by stromal cell-derived factor 1 (SDF-1 or *cxcl12*) and regulates a variety of cellular functions, including chemotaxis, adhesion, haematopoiesis, and organogenesis in many cell types (Moepps et al., 2000). In human, *cxcr4* has six exons (ranging from 31 to 2,668 bp) and two introns (of 2133 and 1616 bp), and it is transcribed into four different tissue-specific variants of different lengths, between 966 and 2668 bp (Thierry-Mieg and Thierry-Mieg, 2006).

The Na^+/Ca^{2+} exchanger gene family (also known as the solute carrier family 8; SLC8) is part of the transporter families of the solute carrier (SLC) gene series, with 43 families and 298 transporter genes currently recognized (Hediger et al., 2004). The solute carrier family 8 includes three members (*slc8a1, scl8a2* and *slc8a3*; also known as *ncx1, ncx2* and *ncx3*, respectively), which likely arose by duplication from a common gene (Lytton, 2007). A fourth member has been recently found in a teleost, but it is absent in both mammals and frogs (*Silurana tropicalis*) (Lytton, 2007; Bowes et al., 2010). The proteins encoded by the members of the SLC8 family exchange extracellular Na^+ by intracellular Ca^{2+} efflux and thus contribute to intracellular Ca^{2+} homeostasis, a function of uttermost importance in regulating cardiac contractility (Quednau et al., 2004). The gene *slc8a1* is expressed ubiquitously while the *slc8a2* and *slc8a3* genes are limited to brain and skeletal muscle. (Quednau et al., 2004). The primary transcript of the *slc8a1* gene is alternatively spliced in a tissue-specific manner, and its expression is regulated by three independent promoters that act on heart, kidney, and all other tissues (Lytton, 2007). In contrast to *slc8a1*, other members of the SLC8 family have a much more restricted pattern of expression: *slc8a2* is abundant in the brain, and *slc8a3* is expressed selectively in skeletal muscle and at lower levels in some brain regions (Quednau et al., 2004; Lytton, 2007). All three members of the SLC8 family display a high degree of sequence identity throughout the length of the protein (Lytton, 2007). In human, the *slc8a1* gene has 11 exons (ranging from 15 to > 6,000 bp), which are separated by 10 introns (the whole region covers more than 340 Kb), and can produce eight different alternative transcripts (Thierry-Mieg and Thierry-Mieg, 2006). The *scl8a3* gene has eight exons (18-2107 bp) and seven introns (882 bp to > 100 Kb) (>144 Kb in total), and can be alternatively spliced to generate up to 10 tissue-specific transcripts (Thierry-Mieg and Thierry-Mieg, 2006).

Rhodopsin (*rho*) belongs to the G-protein-coupled receptor family, whose members are involved in the sense of molecules outside the cell and activation of internal signal transduction pathways (Liang et al., 2003). The protein product encoded by the *rho* gene is a pigment in the retina which is responsible both for the perception of light and the formation of photoreceptor cells (Whitaker

and Knox, 2004). Phototransduction requires the coordinated expression of many genes, including the visual pigments that absorb light (rhodopsin), enzymes involved in the cyclic guanosine-5'-triphosphate (cGMP) cascade that regulates the process, ion channels plus multiple regulatory and structural proteins (Whitaker and Knox, 2004). The *rho* gene has a single copy in the genome and it is composed of five exons (spanning 3.5 Kb of genomic DNA in *Xenopus laevis*; Batni et al., 1996).

Histones are a group of highly conserved small basic proteins that are involved in the packaging and organization of nuclear DNA (Alberts et al., 2002). The *h3a* gene encodes for one of the four core histones (H2A, H2B, H3, and H4), which form a tetramer where the DNA twists around to form the nucleosome (Alberts et al., 2002). The genes encoding for H3 and H4 are 10-fold less divergent than those that encode for H2A and H2B, and core histones that form dimers in the nucleosome (H2A/ H2B and H3/ H4) appear to have co-evolved (Thatcher and Gorovsky, 1994). The *h3a* gene is found in multiple copies in the genome (van Dongen et al., 1981; Turner et al., 1988), and thus, the amplification of different copy variants in different species might potentially represent a problem for phylogenetic inference (Colgan et al., 2000). However, within-species variation among tandem-repeated histone families is unexpectedly low, probably produced by homologous recombination (Maxson et al., 1983). All genes of the histone family appear organized into tandems, whose total length in *Xenopus laevis* is approximately 8.9 Kb (Maxson et al., 1983).

Proopiomelanocortin (*pomc*) is the common precursor protein for a number of peptide hormones and neuropeptides, such as melanotropin (which controls skin darkening), corticotropin and endorphin (Deen et al., 1991). The *pomc* gene is predominantly expressed in the pituitary gland, where the precursor protein is processed into different hormones in each parts of the gland (Deen et al., 1991). In both frogs and mammals, the pomc gene is composed of three exons separated by two long introns, and exon 3 is by far the largest (with a total length of 1001 bp) (Deen et al., 1991). A study comparing different genes showed that *pomc* has a good phylogenetic performance for intermediate levels of divergence (50-300 mya) (Graybeal, 1994).

1.7. Challenges of molecular phylogenetics

1.7.1. Taxon and character sampling

The choice of taxa and characters is a fundamental issue in phylogenetics, but at the same time it is complex and context-specific (Cummings and Meyer, 2005; San Mauro et al., 2009; San Mauro et al., 2012). The traditional approach of comparing the benefit of adding more taxa versus more characters arrived at contradicting conclusions (*e.g.,* Graybeal, 1998; Hillis, 1998; Poe and Swofford, 1999; Rosenberg and Kumar, 2001; Rokas and Carroll, 2005).

For example, using empirical data, Rokas and Carroll (2005) found that increasing the number of genes significantly improved phylogenetic accuracy, whereas increasing the number of taxa did not. However, Graybeal (1998), who used simulated data, found that phylogenetic accuracy improved dramatically with the addition of new taxa and much more slowly with the addition of characters. Apparently, there is no simple answer that can fit all phylogenetic problems, although some general conclusions can be derived from studies that have examined this problematic (Cummings and Meyer, 2005).

A general assumption in phylogenetics is that all genes within a genome share the same evolutionary history (that of the organism). However, sampling only a minuscule part of it can produce, at least potentially, a large variance in the obtained results, and evolutionary processes such as hybridization, introgression, gene duplications, polyploidization, incomplete lineage sorting, horizontal gene transfer, etc. can mislead phylogenetic estimates (Cummings and Meyer, 2005). The choice of particular markers has been generally based on the availability of primers for amplification, perceived general utility, and the expansion of previous datasets (Cummings and Meyer, 2005).

Some studies have examined the relative performance of mt genes (Russo et al., 1996; Zardoya and Meyer, 1996; Miya and Nishida, 2000; Mueller, 2006), and/ or compared the utility of nuclear and mt genes (Graybeal, 1994; Groth and Barrowclough, 1999; Springer et al., 2001; Townsend et al., 2008). In addition, simulation studies have been used to explore how rates of molecular evolution influence phylogenetic reconstruction (Yang, 1998b). However, few general conclusions can be made from the above studies, rather than the relative good performance of mt rRNA genes and the poor performance of the mt gene *nad4L*, and more significantly, the importance of using more than a single gene (San Mauro et al., 2009). The use of several markers can reduce both the variance and the effect of such problems (if present) and consequently, more accurately represent the true evolutionary history of organisms (Cummings and Meyer, 2005).

The choice of taxa has also proven very specific to the phylogenetic question at hand (Cummings and Meyer, 2005). Adding more taxa would in principle improve the accuracy and robustness of inferred phylogenies (Hillis, 1996), although the observation that the levels of homoplasy in the dataset augment with increasing taxon sampling (Sanderson and Donoghue, 1989) has challenged the universality of this statement. The implications and generality of this latter observation remain still debated (Cummings and Meyer, 2005). The problem of taxon sampling implies the questions of how many taxa to include, which taxa to represent and how they should be distributed across the phylogeny (San Mauro et al., 2009; San Mauro et al., 2012). Results from different studies have sometimes arrived at contradicting conclusions, and showed that the answer depends on the details of the particular phylogenetic problem (Sanderson and Donoghue, 1989; Graybeal, 1998; Hillis, 1998; Rosenberg and Kumar, 2001).

The addition of taxa to break up long branches is generally accepted to improve phylogenetic accuracy (Graybeal, 1998), although it depends where the added taxa intersect long branches (Poe and Swofford, 1999; Poe, 2003; San Mauro et al., 2009; San Mauro et al., 2012). The addition of taxa that are internal to a monophyletic group, in contrast, generally increases the support (bootstrap or posterior probabilities) for such groups (Cummings and Meyer, 2005). As an alternative to adding taxa, model-based phylogenetic methods can reduce long-branch attraction problems (Swofford et al., 1996), but if the used model is grossly incorrect the addition of taxa might produce the opposite effect (Poe, 2003).

As a conclusion, the best practice in character sampling is context specific (Russo et al., 1996) and contingent upon taxon sampling, method of analysis, and measures of performance (San Mauro et al., 2009). More sophisticated methods for experimental design attempt to choose certain markers based on their phylogenetic informativeness (Townsend, 2007) or estimate the relative performance of genes and addition of new taxa using the expected Fisher information (Goldman, 1998). These two approaches have been successfully applied to real phylogenetic data, respectively, in fungi (Townsend and López-Giráldez, 2010), and caecilians (San Mauro et al., 2009).

1.7.2. Rate variation among sites

The substitution process is not homogeneous throughout nucleotide (and protein) sequences, and the variation in evolutionary rates across sites has been long acknowledged (Fitch and Margoliash, 1967). This variation is produced by different selective constraints acting in each nucleotide (or protein) site owing to the functional and structural requirements of the gene or protein (Yang, 1996). Clear examples are represented by the different substitution rates among (i) codon positions of protein-coding genes, (ii) structural regions in rRNA genes, and (iii) different proteins, which are more or less evolutionarily conserved depending on their biochemical role (Yang, 1996).

Accounting for among-site rate variation has been shown to improve phylogenetic reconstruction both in model-based distance and maximum likelihood methods, and maximum parsimony (Kuhner and Felsenstein, 1994; Tateno et al., 1994; Gaut and Lewis, 1995; Huelsenbeck, 1995b; Yang, 1995), as well as to more accurately estimate evolutionary distances (hence, branch lengths) (Golding, 1983; Gillespie, 1986; Adachi and Hasegawa, 1995), and transition rate bias (Wakeley, 1994; Yang et al., 1995). Several methods have been proposed to account for such rate variation in phylogenetic analyses. A first approach divides sites into several categories of rates, the simplest way to do this being the assumption that a proportion of sites are invariable (I) while others change at a given (constant) rate (Hasegawa et al., 1985; Palumbi, 1989; Reeves, 1992). A second approach assumes that rates over sites are random variables drawn from a continuous distribution; the most-widely used one being the gamma distribution (Yang, 1996).

The gamma (Γ) distribution is defined by the shape parameter α, with mean 1 and variance 1/α, which can adjust the gamma distribution to different levels or rate variation. For $\alpha \leq 1$ the distribution is L-shaped, while $\alpha > 1$ produces a bell-shaped distribution, that becomes a constant-rate model when α approaches the infinity (Yang, 1996). Although the gamma distribution is continuous, it is usually implemented in a discrete manner using several categories of equal probability (commonly four or eight), as it has been shown to produce a good fit the continuous distribution and it is much more efficient computationally (Yang, 1994, 1996).

1.7.3. Rate variation among lineages

The fact that evolutionary rates can differ among different lineages has been acknowledged for a long time (Simpson, 1944), and a large body of literature supports among-lineage rate heterogeneity at the molecular level, both in mt and nuclear genomes (*e.g.,* Wu and Li, 1985; Hasegawa and Kishino, 1989; Gaut et al., 1992; Clegg et al., 1994; Bromham et al., 1996; Soltis et al., 2002; Tamura and Kumar, 2002; Ho and Jermiin, 2004; Hoegg et al., 2004). A motivation to study and quantify variations in evolutionary rates is the understanding of possible underlying mechanisms that could have triggered the acceleration or slowdown of rates in different lineages (Bromham, 2009a). However, uncovering the causes of lineage-specific rate variation can be a difficult task, and previous studies have attempted to explain them through correlation with species body size, generation time, population dynamics or lifestyle (*e.g.,* Smith and Donoghue, 2008; Bromham, 2009a). Molecular evolutionary rates have also been correlated with diversification (Barraclough and Savolainen, 2001; Eo and DeWoody, 2010; Lanfear et al., 2010), but given the multiple factors that shape diversification patterns, the generalization of this correlation is elusive, and moreover, the cause-effect between rates of genome evolution and cladogenesis remain essentially unknown (Bromham, 2009a).

The interest of studying among-lineage rate variation additionally comes from the field of phylogenetics, because the unequal substitution rates among lineages are a well-known source of phylogenetic artefacts (Philippe and Germot, 2000; Rodríguez-Ezpeleta et al., 2007): rapidly evolving lineages may appear closely related regardless of their true evolutionary relationships (long-branch attraction; Felsenstein, 1978a), whereas short branches may also attract to each other because of the "leftover" similarity of symplesiomorphic states that "eroded" away in rapid-evolving lineages (Fuellen et al., 2001). The long-branch attraction effect was initially described for maximum parsimony and compatibility methods (Felsenstein, 1978a), demonstrating that these methods do not always possess the property of consistency. However, it has been demonstrated that other methods of phylogenetic inference (including maximum likelihood and distance methods) are not immune to the adverse effects of long-branch attraction, especially if the assumed model is too simplistic and it ignores among-site rate variation (Huelsenbeck and Hillis, 1993; Huelsenbeck, 1995a; Philippe, 2000; Philippe et al., 2005; Rodríguez-Ezpeleta et al., 2007; Yang and Rannala, 2012).

Among-lineage rate variation is also an important problem for the estimation of divergence times from molecular data (Yoder and Zhang, 2000; Drummond et al., 2006). Given the inability to separate the individual contributions of rate and time to molecular evolution, estimation of divergence times initially relied on the assumption of a global molecular clock (Zuckerkandl and Pauling, 1962, 1965). However, many studies indicated that clock-like evolution might not be a realistic assumption (Wu and Li, 1985; Hasegawa and Kishino, 1989; Gaut et al., 1992; Clegg et al., 1994; Soltis et al., 2002).

The relaxed molecular clock does not assume a global molecular clock and allows rates to vary across the tree, either (i) assuming that evolutionary rates in closely related lineages are similar (autocorrelated) (Sanderson, 1997; Thorne et al., 1998), (ii) or allowing each branch to possess an independent (uncorrelated) rate (Drummond et al., 2006). Autocorrelation of rates might be expected when the largest component of rate variation is due to inherited factors related to life history traits or biochemical mechanisms (Drummond et al., 2006), and it might be reasonable when sequences have evolved neutrally (*i.e.,* mutation rates correlate with substitution rates) (Ho, 2009). However, substitution rates are determined by the complex interplay between mutation rates, selection, and drift (which might differ among lineages), and it is not clear to what extent substitution rates are inheritable (Ho, 2009). Making an erroneous assumption of rate autocorrelation when rates are actually uncorrelated might severely affect the estimated divergence times, especially in those branches where rates accelerated or slowed down. This is because in autocorrelated dating, large differences in rates among adjacent branches are penalized (*i.e.,* receive low probabilities) (Magallón, 2010), particularly if rate changes occur over short time periods (Welch and Bromham, 2005).

1.7.4. Combined datasets and missing data

Currently, multilocus phylogenetic studies are common and prevail over studies using a single marker, but the question stands of how to best analyze them. Traditionally, two approaches have been used: (i) the construction of phylogenetic trees from individual loci that are later summarized into a consensus tree, or (ii) the analysis of a single dataset containing the concatenation of every locus, this latter option being much more widespread among phylogeneticists (Cummings and Meyer, 2005). In concatenated analysis it is commonplace to apply different evolutionary models for each gene (or other partitions) or use model averaging (Cummings and Meyer, 2005). The concatenation approach tries to maximize the number of informative characters from a set of data, and has the ability to extract "hidden support" for nodes that is only apparent when the loci are concatenated (Gatesy and Baker, 2005; Townsend et al., 2011).

In contrast, building consensus trees from phylogenetic analyses of single loci produces information lost, and some (possibly strongly) supported clades in some analyses might be lost in the consensus (Cummings and Meyer, 2005). However, the concatenation approach may sometimes produce misleading results if the underlying gene-trees are very discordant (Edwards et al., 2007), and thus, the study of congruence among loci prior to analyzing the concatenated dataset might be preferable. A recent approach to analyze multilocus data is represented by species tree methods that allow for independent histories of different loci in terms of topologies and branch lengths (Rannala and Yang, 2003; Edwards et al., 2007; Liu and Pearl, 2007; Liu, 2008; Kubatko et al., 2009; Heled and Drummond, 2010). Most species tree methods implement a coalescent model in a Bayesian framework, whose full potential requires the use of intraspecific information (*e.g.,* Recuero et al.; Townsend et al., 2011), although it is not restricted to it (*e.g.,* Bewick et al., 2012).

A recent study (Fisher-Reid and Wiens, 2011) examined the consequences of combining mt and nuclear gene sequences in several datasets, finding that discordance between trees derived from mt and nuclear data was common, but this discordance was typically not strongly supported. In general, congruence between mt and nuclear data was higher on branches that are longer and deeper in the combined-data tree, probably because gene histories are more likely to coalesce (Fisher-Reid and Wiens, 2011). Nuclear genes generally dominate the phylogenetic signal over mt genes at deeper nodes (but not always), as their slower substitution rates make them less likely to be saturated at deeper divergences (Fisher-Reid and Wiens, 2011). However, although mt data is in principle more informative for recent divergences, whether mt data dominated over nuclear genes or not at shallower divergences varied among the case studies (Fisher-Reid and Wiens, 2011). Nevertheless, the above generalities should be taken with caution: both the presence of small or high discordance between mt and nuclear markers, as well as whether a particular node is resolved in favour of mt or nuclear data, depends significantly on various factors, including the genes and taxa sampled, their levels of divergence, evolutionary rates of genes, and the proportion of information to noise in each dataset.

Besides the strategy chosen to combine multilocus data, completeness of matrices is an important issue. Several studies that examined the use of sparse matrices have shown that phylogenetic analyses can be accurate despite a high proportion of missing data, as long as the overall number of characters is large and informative (Wiens, 2003; Driskell et al., 2004; Philippe et al., 2004; Wiens and Moen, 2008; Wiens and Morrill, 2011). In general, adding taxa with missing data to monophyletic groups is less likely to decrease the accuracy of phylogenetic methods, but adding taxa with missing data in situations involving long-branch attraction might mislead phylogenetic accuracy (Cummings and Meyer, 2005).

However, when missing data is non-randomly distributed, the effects of missing data might be quite severe (Simmons, 2012). This is particularly important when model-based phylogenetic methods are used, because extensive missing data can impact the estimation of parameters that are based on summation of all characters (*i.e.,* branch lengths, compositional biases, corrected distances between taxa, etc.), producing a negative impact on the obtained phylogeny (Gatesy et al., 2002). This contrasts with maximum parsimony analyses, where missing data only affect those characters in particular for which information is not present (Gatesy et al., 2002). For this reason, missing data may decrease resolution in parsimony analyses, but it is unlikely to mislead the resulting phylogeny (Huelsenbeck, 1991).

1.7.5. Amino acids versus nucleotides

The evolutionary history behind a protein-coding sequence is only one, and thus, the phylogenetic relationships inferred from it should, in principle, be the same regardless of whether nucleotide- or amino acid-level data are used (Gissi et al., 2006). Nevertheless, several studies have shown that evolutionary histories inferred from a given gene (nucleotides) or its product (amino acids) might differ considerably (*e.g.,* Townsend et al., 2008). This is due to the fact that nucleotides and amino acids posses different attributes and retain phylogenetic signal in a different way (Gissi et al., 2006).

Amino acid data are generally preferred for deeper divergences, as replacement rates of amino acids are slower than nucleotide substitution rates (because only non-synonymous substitutions are reflected in the amino acid level), and thus amino acids are less likely to be saturated when divergent taxa are compared (Simmons et al., 2004). Conversely, the use of nucleotide data is favoured to infer relationships among closely related taxa (Simmons et al., 2002; Simmons et al., 2004). The exclusion of third codon positions is a common practice when analyzing deep phylogenetic divergences because they are more likely to be saturated by synonymous substitutions (Cao et al., 1994) and often show strong compositional biases (Chang and Campbell, 2000; Ho and Jermiin, 2004).

A second important difference is that amino acids posses a larger character-state space compared to nucleotides (20 vs. 4, respectively), making them less likely to convergence (Simmons et al., 2004). Third, nucleotides have a threefold advantage in number of characters (or double if third codon positions are excluded) (Simmons et al., 2004). A fourth difference is that amino acids are less likely to show compositional biases (Loomis and Smith, 1990; Lockhart et al., 1992; Hashimoto et al., 1995; Simmons et al., 2004), even though they are not completely immune to this effect (Foster et al., 1997; Singer and Hickey, 2000). Phylogenetic analyses of nucleotides use 4 x 4 substitution matrices, which are more tractable than 20 x 20 amino acid matrices, making analyses noticeably faster (Reeves, 1992), even though the difference in computational power might only be appreciable when very large data matrices are being analyzed.

The choice of nucleotides or amino acids might affect both the inferred topologies and statistical support (Gissi et al., 2006). However, studies comparing the relative performance of either nucleotides or amino acids in different phylogenetic frameworks seem to arrive at contrasting conclusions, suggesting that the problem might be context-dependent. Simmons et al. (2002; 2004) found that nucleotides outperformed amino acids both in topological resolution and branch support in recovering deep phylogenetic relationships. Simulation studies by Hall (2005) failed to make any generalization on whether the performance of different methods of phylogenetic inference is equally affected by the differences in phylogenetic signal of amino acids or nucleotides, but indicated that Bayesian inference analyses using nucleotides were the most accurate, followed by the maximum likelihood tree based on nucleotides, and the maximum parsimony tree using protein data.

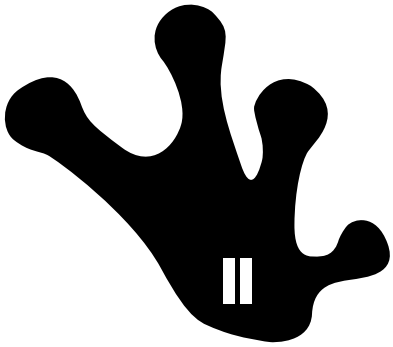
In principle, one would expect model-based phylogenetic methods to be less influenced by the choice of either nucleotides or amino acids, specially if data is appropriately partitioned, given that evolutionary models can account for compositional and among-site rate heterogeneity, and ultimately maximize the recovery of phylogenetic signal from both types of data (Gissi et al., 2006). Interestingly, Townsend et al. (2008) studied the phylogenetic informativeness of nucleotide and amino acid data from 11 mt and nuclear genes of vertebrates, and arrived to the conclusion that nucleotides have greater phylogenetic signal than amino acids due to their threefold greater representation, encouraging their use over all time scales where they are not subjected to strong convergence biases. These authors also found that nucleotide-based analyses produced only moderately higher support values, and that amino acid data showed very low levels of noise, producing higher support values even when the net amount of signal was low (Townsend et al., 2008).

1.7.6. Paralogy

The concepts of orthology and paralogy are hierarchical-dependent. For example, in the case of genes of the globin family that arose by duplication in chordates, several paralogous genes exist within a species, but each paralogous gene possesses its orthologous copy in other chordate species. In principle, the inference of phylogenetic relationships among organisms should only be based in the comparison orthologous loci (Swofford et al., 1996). However, gene duplications have been common throughout vertebrate evolution, and thus, phylogenetic inference from nuclear genes is challenged by the probability of sampling paralogous genes (Cotton and Page, 2002). This is the case of both paralogous genes from a gene family (*e.g.,* heat shock 70 genes; Martin and Burg, 2002), or multicopy genes that appear several times in the nuclear genome (for example, nuclear rRNA genes and their internal transcribed spacers; Álvarez and Wendel, 2003). Furthermore, paralogous genes may sometimes seem orthologous due to a differential lineage-specific gene loss; that is, when a different paralogous copy is lost in two taxa being compared (Koonin, 2005).

Paralogy represents a real problem for phylogenetic inference using nuclear genes, as detecting paralogous loci might be a difficult task (Cotton and Page, 2002; Koonin, 2005). However, orthology assessment is facilitated by the availability of fully sequenced nuclear genomes and the use of phylogenetic tools (Gabaldón, 2008). Paralogy has long been considered a minor problem in the mt genome, and the direct orthology of mt genes has been pointed out (Funk and Omland, 2003). Although this assertion remains largely valid, it should be kept in mind that the PCR amplification with mt-specific primers from whole-genomic DNA might bring up two problems: nuclear pseudogenes and paternal inheritance (Funk and Omland, 2003).

Nuclear pseudogenes are segments of the mt genome that were transferred to the nucleus, where they become functionless (Collura and Stewart, 1995; Bensasson et al., 2001). Pseudogenes might be amplified by mistake, but they can be usually detected by their unusual patterns of molecular evolution: presence of indels, frame shifts, stop codons, elevated frequencies of non-synonymous substitutions, different compositional biases, or slowed substitution rates (Funk and Omland, 2003). Unrecognized biparental inheritance of mt genomes might also mislead phylogenetic inference, but this has been shown to be very restricted to bivalve molluscs (Rawson and Hilbish, 1995; Breton et al., 2007), and phylogenetic relationships among affected taxa can be successfully reconstructed once biparental inheritance is acknowledged (*e.g.,* Doucet-Beaupré et al., 2010).



OBJECTIVES

OBJECTIVES

This Ph. D. thesis aims to study the phylogenetic relationships among major lineages of frogs and resolve a number of controversial issues by using molecular information from complete mt genomes and nine partial nuclear genes. The thesis is structured into three sections.

The specific objectives of the work are the following:

1. To study the relative phylogenetic position of the basal genera *Ascaphus* and *Leiopelma* within Anura, in order to discriminate among competing hypotheses regarding the root of the frog tree of life (Section I).

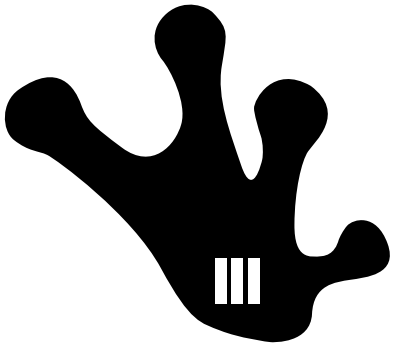
2. To determine the phylogenetic position of the clade Pipoidea within Anura and the intergenetic relationships within the family Pipidae (Section II).

3. To study the evolution of sound production mechanism in Pipidae in the context of a robust phylogenetic framework by means of both behavioural and anatomical observations (Section II).

4. To gain insight into the controversial phylogenetic relationships among basal families within Neobatrachia (Section III).

5. To study the neobatrachian-specific substitution rate acceleration both in mt and nuclear genes, precisely determining the time frame and possible causes of its origin (Section III).

6. To determine the gene order of all newly sequenced mt genomes, in order to study possible gene rearrangements and mechanisms, with emphasis on the origin of the neobatrachian-specific gene order (Sections I, II and III).



MATERIALS AND METHODS

3.1. Taxon and character sampling

In order to address the different questions proposed in this thesis, two main types of genetic markers were used: (i) complete mt genome sequences, and (ii) partial sequences of nine nuclear protein-coding genes: *bdnf*, exon 2 of *cxcr4*, *h3a*, *pomc*, *rag1, rag2*, exon 1 of *rho*, exon 2 of *slc8a1* and *slc8a3*. Newly generated data from complete mt genomes was complemented with previously available data from GenBank (Table 1.2). For the nuclear dataset, a smaller set of taxa was chosen among those species with available mt genomes, and representing major lineages of frogs. New data from nuclear loci was generated in order to complement previously available sequences from GenBank, creating an almost complete data matrix (*ca.* 90% complete). In a few cases, and only in the nuclear dataset, chimerical sequences were constructed by merging sequences from different species of the same genera, for which strong evidence exist of being monophyletic. Taxon sampling was slightly different for each particular question addressed, and the new sequence data generated during this Ph. D. thesis has been progressively incorporated into subsequent studies as they became available. The taxon sampling strategies are explained below, and detailed information of species names, genetic markers, corresponding GenBank accession numbers, tissue and/ or DNA vouchers, and information of collection localities is given in Appendix I.

In the first study, the complete sequence of the mt genome of *Leiopelma archeyi* was determined. Taxon sampling among available species in GenBank was designed to represent main lineages within extant frogs: *Alytes obstetricans, Ascaphus truei, Bombina orientalis, Discoglossus galganoi, Pelobates cultripes,* and *Silurana tropicalis* among non-neobatrachians, and two species from each of the two highly diverse lineages within Neobatrachia, Hyloides (*Duttaphrynus melanostictus* and *Hyla chinensis*) and Ranoides (F*ejervarya limnocharis* and *Pelophylax nigromaculatus*) (*sensu* Frost et al., 2006). Four salamander species were used as outgroups: *Ambystoma mexicanum, Andrias davidianus, Lyciasalamandra atifi,* and *Ranodon sibiricus.*

In the second study, the mt genomes of four pipoids were sequenced anew (*Rhinophrynus dorsalis, Pipa carvalhoi, Hymenochirus boettgeri* and *Pseudhymenochirus merlini*) and the relatively old available sequence of *Xenopus laevis* (Roe et al., 1985) was replaced by a newly determined one from a specimen with reliable locality data. The mt dataset included the new mt genomes along with those of all other frogs available in GenBank (in 2009); and it was reduced to represent main major lineages of frogs in the nuclear dataset.

In the third study, ingroup species were chosen in order to represent main lineages within Anura, and they were complemented with seven new complete mt sequences from one pelobatoid (*Pelodytes punctatus*; Pelodytidae) and the following neobatrachians: *Heleophryne regis* (Heleophrynidae), *Lechriodus melanopyga* (Limnodynastidae), *Calyptocephalella gayi* (Calyptocephalellidae), *Telmatobius bolivianus* (Ceratophryidae), and *Sooglossus thomasseti* (Sooglossidae). The nearly complete sequence of another Sooglossidae, *Sooglossus sechellensis*, was also determined. New sequence data of nine nuclear genes for the aforementioned species was generated in order to complete the nuclear dataset. Outgroup sampling was expanded for the estimation of divergence times in order to include additional calibration points, including three salamanders (*Siren intermedia, Andrias davidianus* and *Batrachuperus pinchonii*), three caecilians (*Rhinatrema bivittatum, Ichthyophis glutinosus* and *Typhlonectes natans*), a lizard (*Iguana iguana*), a bird (*Gallus gallus*) and a mammal (human).

3.2. Laboratory procedures

Total DNA was purified from ethanol-preserved tissue by proteinase k digestion and either phenol-chloroform extraction (Sambrook et al., 1989) or salt-extraction (Bruford et al., 1992), followed by ethanol purification (Sambrook et al., 1989).

Mitochondrial genomes were amplified in several overlapping fragments by PCR using the primers and cycling conditions reported in San Mauro et al. (2004b). Due to the presence of gene rearrangements, the mt genomes of neobatrachian species were partially amplified (*ca.* 8-10 Kb, from 5'–*cox3* to 3'–*trnF*) using the primers and conditions reported in Kurabayashi and Sumida (2009). Specific primers were also designed to amplify certain fragments when general primers did not work (mainly for long and complex control regions that typically contain repetitive elements, or due to the gene rearrangement found in *Leiopelma archeyi;* see Reults 4.1.1), and to obtain the full sequence of long PCR fragments by primer walking. Partial sequences of nuclear genes were amplified using the primers and conditions reported in the literature: *rag1* (San Mauro et al., 2004b); *rag2* (Venkatesh et al., 2001; Hoegg et al., 2004); *slc8a1* (Roelants and Bossuyt, 2005); *bdnf* and *pomc* (Vieites et al., 2007); *rho* (Hoegg et al., 2004); and *h3a* (Colgan et al., 2000).

In all cases, PCR cycling conditions were experimentally adjusted from those reported in the original publications. Fragments up to 1,500 bp were amplified in 25 µl reactions containing 2.5 µl 10x standard reaction buffer, 1.5 µl $MgCl_2$ 25 mM, 0.5 µl of dNTPs (desoxiribonucleoitde triphosphates) mixture 10 µM (2.5 mM each), 0.5 µl of each primer at 10 µM, 0.5 µl of total DNA (10-100 ng), and 0.2 µl of *Taq* DNA polymerase (5PRIME GmbH, Hamburg, Germany). Cycling conditions were as follows: an initial denaturing step at 94 ºC for 5 min; followed by 40-45 cycles

of denaturing at 94 ºC for 60 s, annealing at 42-54 ºC for 60 s, and extending at 72 ºC for 90 s; and a final extension step of 72 ºC for 5 min.

Longer PCR fragments (1.5-6 Kb) were amplified using LA *Taq* polymerase (TaKaRa Bio Inc., Otsu, Shiga, Japan), in 25 µl reactions containing 2.5 µl 10x LA PCR buffer II (Mg^{2+} plus), 4 µl of dNTPs mixture 2.5 µM (2.5 mM each), 0.5 µl of each primer at 10 µM, 0.5 µl of total DNA (10-100ng), and 0.25 µl of LA *Taq* DNA polymerase. Cycling conditions were as follows: an initial denaturing step at 98 ºC for 30 s; followed by 40-45 cycles of denaturing at 98 ºC for 10 s, annealing at 42-68 ºC for 60s, and extending at 68 ºC for 2-4 min; and a final extension step of 68 ºC for 10 min. The PCR conditions for the partial amplification of neobatrachian mt genomes strictly followed Kurabayashi and Sumida (2009).

PCR amplicons were purified by ethanol precipitation (Sambrook et al., 1989) or directly from electrophoresis gels using the Speedtools PCR clean-up kit (Biotools B&M Labs. S.A., Madrid, Spain). The long-PCR products containing the control region of *L. melanopyga* and *T. bolivianus* were digested with *Pst*I at 37 ºC for 4 hours, obtaining two fragments from each of the original amplicons. These four fragments, as well as all other PCR products containing the control regions of the remaining species were cloned into pGEM-T vectors (Promega, Madison, WI, USA). PCR fragments and positive recombinant clones were cycle-sequenced with the ABI Prism BigDye Terminator v. 3.1 cycle sequencing ready reaction kit (Applied Biosystems, Foster City, CA, USA) using PCR and M13 universal primers, and following manufacturer's instructions. Cycle sequencing products were run on ABI Prism 3700 and 3130xl DNA Analyzers (Applied Biosystems, Foster City, CA, USA).

3.3. Annotation of mitochondrial genomes

The new mt sequences were annotated by comparison with other reported vertebrate mt genomes using the web-based annotation tool DOGMA (Wyman et al., 2004). In this pipeline, all genes are identified based on conservation of sequence similarity to other genes in a custom database using BLAST (Altschul et al., 1990). For protein-coding genes, the mt genome is translated into all six reading frames (using the vertebrate mt genetic code) and queried against a custom amino acid database. Ribosomal and transfer RNA genes are queried against a nucleotide sequence database. Because tRNA genes can be very divergent among animal mt genomes, sequence similarity is not sufficient to provide a confident identification (Wyman et al., 2004), and therefore, DOGMA implements the COVE program (Eddy and Durbin, 1994), which uses hidden Markov models to predict the conservation of base pairings in the cloverleaf secondary structure. The putative secondary structure of tRNA genes is drawn with a custom program in DOGMA (Wyman et al., 2004).

Initial and final position for both protein-coding genes (start and stop codons) and rRNA genes must be manually selected. In all cases, initial and final positions of genes were selected to maximize similarity to other annotated mt genomes and to avoid or minimize gene overlapping (Boore et al., 2005; Sheffield et al., 2010). Selecting initial and final positions of rRNA genes is particularly difficult, and thus, they were assumed to extend to the boundaries of adjacent genes (Boore et al., 2005). All non-coding regions were easily recognizable as they lay between the boundaries of other genes, and furthermore, the location of the two main non-coding regions is fairly conserved among vertebrates. The control region is typically found between the *trnP* and *trnF* genes, and features such as putative CSB or TAS were identified through BLAST searches. The origin of replication of the light strand is typically found between the *trnN* and *trnC* genes, within the *WANCY* region, and it was easily recognized by its secondary structure. Other intergenic spacers are typically absent in mt genomes, and if present, they tend to be very short (e.g., 15-47 bp in *Leiopelma archeyi*).

3.4. Phylogenetic analyses

3.4.1. Multiple sequence alignment and removal of poorly aligned positions

Sequence alignment is a prerequisite for inferring phylogenies and studying evolutionary patterns and rates of sequence divergence. Sequence alignment represents a hypothesis about positional homology of nucleotide or amino acid residues from different organisms, and it is operationally solved by means of mathematical algorithms.

Multiple sequence alignments are generally calculated using heuristics as an alternative to the dynamic programming, which is computationally very expensive (Higgins and Lemey, 2009). The software MAFFT is based on the fast Fourier transform for the rapid identification of homologous segments, and uses either the progressive or the iterative refinement heuristic methods along with an improved scoring matrix (Katoh et al., 2002). In the progressive alignment method, sequences are gradually aligned following the branching order of a (distance-based) guide tree (Higgins and Lemey, 2009), and MAFFT implements the FFT-NS-2 algorithm, which repeats the above procedure by calculating pairwise distances from the alignments and provides a second, refined alignment (Katoh and Toh, 2008).

In the iterative refinement method, an alignment calculated with the progressive method is partitioned into two groups, which are then realigned and merged into a new alignment; this process is done iteratively until no more improvements are made (Katoh and Toh, 2008). Objective functions are used to measure the "goodness" of a given alignment; in the case of G-INS-i, L-INS-i

and E-INS-i methods, both weighted sum-of-pairs and consistency (between pairwise and multiple alignments) objective functions are used (Katoh and Toh, 2008). The above three methods are slow and accurate, but differ in the way local pairwise alignment are done; for example, the method L-INS-i is particularly suited for sequences with one conserved domain and long gaps (Katoh and Toh, 2008)

After multiple sequence alignments are generated, poorly aligned positions should be removed because they may not be homologous or may have been saturated by multiple substitutions, making positional homology of such positions unreliable. The removal of poorly aligned positions has been shown to improve the results of phylogenetic inference (Castresana, 2000; Talavera and Castresana, 2007). The software Gblocks (Castresana, 2000) is an automatic method to remove these problematic regions, and it is based on the selection of blocks of positions that fulfil a set of requirements with respect to the number of contiguous conserved positions, lack of gaps, and high conservation of flanking positions.

TranslatorX (Abascal et al., 2010) is a web-based tool designed to align protein-coding nucleotide sequences based on their corresponding amino acid translations. Hence, codons are preserved in the nucleotide alignments, maximizing positional homology of codon positions and ensuring that no frame shifts are present (which is essential for codon-based analyses). TranslatorX implements several alignment methods (in the present case MAFFT was used) to construct amino acid alignments. It also implements Gblocks (Castresana, 2000) to remove poorly aligned amino acid positions prior to creating the protein-based nucleotide alignment.

In the first study, the 12 protein-coding genes encoded by the heavy strand of the mt genome were used. The *nad6* gene (encoded by the light strand) was excluded to avoid possible biases due to the known difference in base composition between the two strands (Reyes et al., 1998), which could translate into amino acid compositional differences (Foster et al., 1997; Singer and Hickey, 2000) and introduce additional biases in the phylogenetic inference. Amino acid sequences from individual genes were manually aligned against a previous database (San Mauro et al., 2004a; Gissi et al., 2006) and all 12 genes were concatenated into a single dataset. Gblocks v. 0.91b (Castresana, 2000) was used to exclude alignment ambiguities, using default parameter settings.

In the second and third studies, individual protein-coding genes were aligned with TranslatorX (Gblocks used the following settings: smaller final blocks, gap positions within the final blocks and less strict flanking positions). Transfer RNA genes were aligned manually based on their putative secondary structure, and concatenated into a single dataset, whereas rRNA genes were aligned with MAFFT L-INS-i (Katoh et al., 2002) and corrected by eye for any obvious misalignment. Ambiguously aligned positions in both rRNA and tRNA alignments were also excluded with Gblocks v. 0.19b (Castresana, 2000) with the same options used for protein-coding genes.

3.4.2. Saturation, removal of fast-evolving sites, and concatenation

Plots of pairwise uncorrected *p* versus patristic distances measured as maximum likelihood distances (using the general time-reversible [GTR+I+ Γ] model in PAUP* v. 4.0b10; Swofford, 2003) indicated the presence of saturation at third codon positions of mt protein-coding genes (Fig. 3.1) and thus, they were excluded from the final nucleotide datasets.

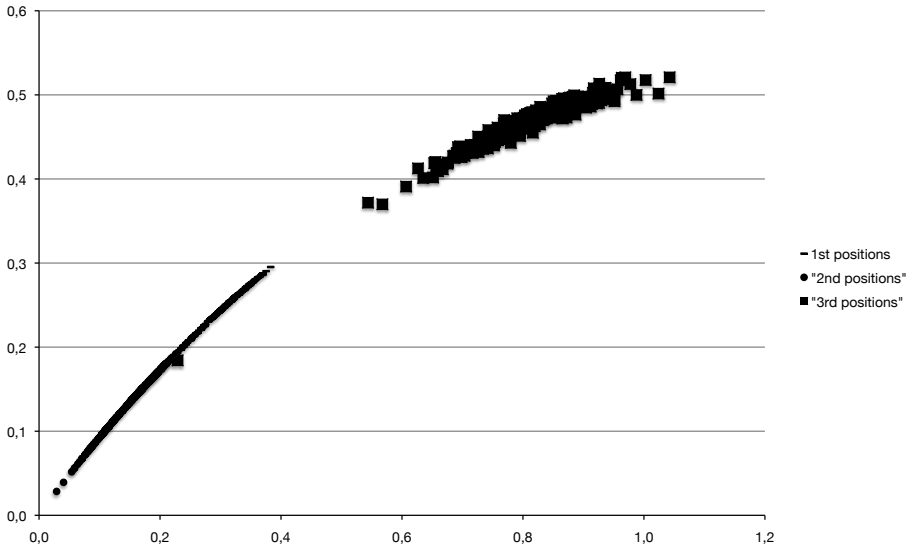


Fig. 3.1. Pairwise uncorrected *p* versus patristic maximum likelihood distances (assuming a GTR+I+Γ model) in mt protein-coding genes. Data is from the third and most complete study. Note that third codon positions are saturated.

In the first study that aimed to establish the phylogenetic position of the basal genera *Ascaphus* and *Leiopelma*, fast-evolving amino acid positions were removed from the initial dataset in order to avoid the prejudicial effect of long-branch attraction between highly divergent neobatrachians and distantly related salamanders, as reported in previous mitogenomic studies (San Mauro et al., 2004a; Gissi et al., 2006). The selection of fast evolving sites was done using the Γ distribution of among-site rate heterogeneity as a selective criterion in a maximum likelihood framework. PAML v. 3.15 (Yang, 1997) was used to assign each site of the alignment into one of the eight categories of the Γ distribution, and all sites assigned to the two fastest evolving categories (with replacement rates of 4.48 and 1.76 replacements · site^{-1}, respectively) were manually removed from the alignment. The six remaining slower evolving categories (sites retained in the final data set) had rates of evolution of less than 0.90 substitutions · site^{-1}. The final dataset consisted in a single partition with 2,498 aligned amino acid positions (Table 3.1).

In the second study, the resulting single-gene alignments were merged into three datasets, containing (i) all 37 mt genes (final length of 11,131 bp), (ii) all nine nuclear genes (final length of 7,107 bp), and (iii) a combination of all mt and nuclear genes (final length of 18,238 bp). Three different partitioning schemes were used to analyse all three datasets: (i) a single partition, (ii) partition by gene, and (iii) partition by gene and codon position (Table 3.1). Additionally, nuclear single-gene alignments were analysed to gain insight into the congruence among these markers.

In the third study, original single-gene alignments and the concatenation of all mt and all nuclear genes were used to examine known biases in substitution rates by using (i) relative-rate tests, (ii) branch length measurements, and (iii) differences in selection coefficients at the molecular level (see below). Preliminary phylogenetic analyses of separate mt and nuclear datasets suggested a long-branch attraction artefact (see Results 4.3.2 and Fig. 4.11), and thus, single-gene alignments were reduced to retain only the more conserved positions. For protein-coding genes, Gblocks was employed, using default parameters (Castresana, 2000), and first and third codon positions of mt genes, and third codon positions of nuclear genes were further excluded. For rRNA and tRNA genes the following settings were used in Gblocks: minimum number of sequences for a conserved position 31, minimum number of sequences for a flanking position 36, maximum number of contiguous non-conserved positions 5, minimum length of a block 10, allowed gap positions with half. The resulting combined reduced dataset (Table 3.1) was divided into 5 partitions: (i) second codon positions of all mt protein-coding genes, (ii) mt rRNA genes, (iii) mt tRNA genes, (iv) first codon positions of all nuclear genes, and (iv) second codon positions of all nuclear genes.

3.4.3. Model selection and hypothesis testing

In model-based phylogenetic methods (maximum likelihood and Bayesian inference) the use of an appropriate model is essential to obtain accurate estimates of the phylogeny, the measures of confidence, and for phylogenetic hypothesis testing (Tamura, 1994; Zhang, 1999; Buckley and Cunningham, 2002; Lemmon and Moriarty, 2004; Posada and Buckley, 2004). Model selection strategies use the likelihood function as a measure of model fit (usually the maximized log likelihood; Posada, 2009) and then sort the different models according to their fit to the data.

A widely accepted statistic for testing the relative fit of models is the likelihood ratio test (LRT) statistic, which is χ^2–distributed for nested models (Goldman, 1993). However, the use of the hierarchical likelihood ratio test for model selection is discouraged (Posada and Buckley, 2004) in favour of other alternatives. The AIC (Akaike, 1973) is an asymptotically unbiased estimator of the Kullback-Leibler information quantity (Kullback and Leibler, 1951), and represents the amount of information lost when using a certain model; hence the model with the smallest AIC score is the model that best fits the data (Posada, 2009). A Bayesian alternative to AIC is the BIC

(Schwarz, 1978). Both AIC and BIC can be used to compare both nested and non-nested models (Posada, 2009). Additionally, the difference between AIC (ΔAIC) or BIC scores provide a sense of the confidence of the tested models, allowing an easy and quick comparison and ranking of all candidate models (Posada, 2009).

Several available software programs implement the above strategies for selection of best-fit models for both nucleotide and amino acid data. Modeltest v. 3.7 (Posada and Crandall, 1998) and jModeltest v. 0.1.1 (Posada, 2008) were used to select the most appropriate models for maximum likelihood analyses, while MrModeltest (by J. A. A. Nylander; http://www.abc.se/~nylander/) was used for Bayesian inference, as it is a reduced version of Modeltest, testing only the models implemented in MrBayes (Ronquist and Huelsenbeck, 2003). Prottest (Abascal et al., 2005) was used to select the best-fit amino acid replacement model for amino acid-based analyses in both maximum likelihood and Bayesian inference analyses. The most appropriate models of evolution were chosen for each partition using the AIC criterion and then accordingly applied into phylogenetic inference software.

The above strategies of model selection are not restricted to the choice of appropriate models of sequences evolution; indeed, they are widely used in hypothesis testing in the phylogenetic context, including the comparison of the relative fit of alternative models that assume different rates of silent and replacement substitutions (Yang, 1998a; Yang and Nielsen, 1998).

Table 3.1. Summary of datasets used in phylogenetic analyses

Dataset	No. Positions	Outgroup (No.)	No. ingroup species	Partitions (No.)	Results
Study I					
Mitochondrial	2,498	salamanders (4)	11	single	Fig. 4.2
Study II					
Mitochondrial	11,131	*Leiopelma* + *Ascaphus*	35	single, by gene (16), by gene and codon position (42)	Fig. 4.3
Nuclear	7,107	*Leiopelma* + *Ascaphus*	19	single, by gene (9), by gene and codon position (27)	Fig. 4.3
Combined dataset	18,238	*Leiopelma* + *Ascaphus*	19	single, by gene (25), by gene and codon position (69)	Fig. 4.3
Single nuclear genes	507-1,512	*Leiopelma* + *Ascaphus*	15-19	single	Table 4.4
Study III					
Mitochondrial	14,463	*Leiopelma* + *Ascaphus*	26	by gene (16)	Fig. 4.11
Nuclear	7,083	*Leiopelma* + *Ascaphus*	26	by gene (9)	Fig. 4.11
Combined reduced	11,136	Leiopelma + Ascaphus	26	genes and codon positions (5)	Fig. 4.11 Fig. 4.12

3.4.4. Maximum likelihood

The maximum likelihood analyses were conducted with RAxML v. 7.0.4 (Stamatakis, 2006) using the rapid hill-climbing algorithm (Stamatakis et al., 2007) to compute 100 distinct maximum likelihood trees starting from 100 distinct randomized maximum parsimony trees. RAxML is a software for rapid an efficient maximum likelihood analyses, and uses a tree search heuristic method called lazy subtree rearrangement, which is a variant of SPR (subtree prunning and regrafting) (Schmidt and von Haeseler, 2009). Table 3.1 shows a summary of datasets used in phylogenetic inference in all three studies.

3.4.5. Bayesian inference

For Bayesian inference, MrBayes v. 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) was used, running four simultaneous Markov chains for 10-20 million generations, sampling every 1000 generations, and discarding the first 1-2.5 million generations as burnin to prevent sampling before reaching stationarity (details are given in Table 3.2). Two independent Bayesian inference runs were performed to confirm the adequate mixing of the Markov chains and to increase the chance of detecting failure to converge. Convergence of Markov chains was checked a posteriori by plots of log maximum likelihood scores and low standard deviation of split frequencies, as well as using the convergence diagnostics implemented in the online tool AWTY (Nylander et al., 2008) and the program Tracer v. 1.5 (Rambaut and Drummond, 2009).

Table 3.2. Summary settings used in the different Markov Chain Monte Carlo searches.

Study	No. Runs	No. Chains (per run)	No. generations (per chain)	Chain thinning	Burnin
Study I	2	4	10 millions	1000	1 million
Study II	2	4	10 millions	1000	2.5 millions
Study III	2	4	20 millions	1000	1 million

3.4.6. Node support

Maximum likelihood analyses produce only point estimates of the phylogeny without a measure of their confidence, and for this reason, several methods have been developed to quantify the robustness of the obtained phylogeny (Felsenstein, 1985b; Strimmer and von Haeseler, 1996; Anisimova and Gascuel, 2006). The most widely used approach to assess the reliability of the obtained phylogenetic trees is the non-parametric bootstrapping (Efron, 1979; Felsenstein,

1985b). Non-parametric bootstrapping is a resampling technique in which bootstrap samples of the same size as the original data are created by randomly drawing *n* columns with resampling from the original alignment (Felsenstein, 1985b). Each bootstrap sample is used to reconstruct a phylogenetic tree, and the variation among the resulting trees is taken as a measure of confidence of the results from the original dataset. This can be done with a majority-rule consensus tree, and node support is usually expressed in per cent.

Although the exact interpretation of the statistical significance of bootstrap proportions is elusive (Berry and Gascuel, 1996; Susko, 2010), a value of 70% or greater in a node is considered indicative of substantial confidence for that particular relationship (Hillis and Bull, 1993,). Moreover, it is not clear how many replicates are required to provide reliable support values (but see Hedges, 1992; Pattengale et al., 2009), although 1,000 replicates are generally considered an adequate conservative estimate (Pattengale et al., 2009).

Therefore, 1,000 bootstrap replicates were performed in RAxML v. 7.0.4 (Stamatakis, 2006). In the Bayesian framework, node support is typically assessed by summarizing topology posteriors and visualized it in a majority rule consensus tree (Ronquist et al., 2009). Some studies have cautioned that posterior probability measures can be reflect overconfidence (Suzuki et al., 2002), and values higher than 0.95 are generally considered to provide strong support for a particular node (Alfaro et al., 2003; Erixon et al., 2003; Huelsenbeck and Rannala, 2004). In Bayesian inference analyses, support for internal branches was evaluated by posterior probabilities given by MrBayes v. 3.1.2 (Ronquist and Huelsenbeck, 2003).

3.4.7. Testing tree topologies

Several different techniques can be used to compare contradicting (hence, non-nested) tree topologies, but the most commonly used methods are based on the comparison of the likelihood values from different topologies (Schmidt, 2009). The approximately unbiased test (AU; Shimodaira, 2002) is less biased than other tree selection techniques (Shimodaira, 2002), and it uses a multiscale bootstrap approach to control for type 1 errors, while reducing the excessively conservative tree selection biases of other non-parametric tests, such as the Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999), often accused of being very conservative (Goldman et al., 2000; Shimodaira, 2002). The multiscale bootstrap procedure consists on generating several sets of bootstrap replicates with changing sequence lengths (they may differ from that of the original data). Then, the number of times a given hypothesis is supported by the replicates is counted for each set to obtain bootstrap proportion values for different sequence lengths. Finally, the approximately unbiased p value is calculated from the change in the bootstrap proportion values along the changing sequence length (Shimodaira, 2002).

The AU test was used to evaluate alternative phylogenetic hypotheses (taken from the literature) to those supported by our data. The program Consel v. 0.1i (Shimodaira and Hasegawa, 2001) was used to generate p values from AU tests, while site-wise log-likelihoods were calculated with PAML v. 3.15 (Study I; under the mtREV+Γ model) or RAxML v. 7.0.4 (Study II; under independent GTR+I+Γ models for each gene partition) and one million multiscale bootstrap replicates in both cases.

3.5. Estimation of divergence times

The information contained in sequence data can be used both to infer the phylogenetic relationships among organisms, as well as to estimate the absolute timing of their divergence. The now popular relaxed clock methods go beyond the original proposition of the molecular clock (Zuckerkandl and Pauling, 1965), accounting for the rate variation observed among different lineages, thus allowing more accurate time estimates (Hedges and Kumar, 2009). The most commonly used methods implementing relaxed molecular clocks are penalized likelihood (Sanderson, 1997, 2002), Bayesian rate autocorrelation dating (Thorne et al., 1998; Kishino et al., 2001; Thorne and Kishino, 2002) and Bayesian uncorrelated relaxed clocks (Drummond et al., 2006; Drummond and Rambaut, 2007). The first two methods assume that evolutionary rates in closely related lineages are similar (autocorrelated), while the third method does not. The two Bayesian methods permit the use of multiple loci and partitioned data and are highly parametric, allowing the extraction of more information about the evolutionary processes from sequence data (San Mauro and Agorreta, 2010).

In the Bayesian uncorrelated relaxed clock, evolutionary rates are drawn independently and identically in each branch from an underlying lognormal distribution (Drummond et al., 2006). The choice of appropriate calibrations is one of the most important decisions to be made in molecular dating (Graur and Martin, 2004; Hedges and Kumar, 2004; Donoghue and Benton, 2007; Ho, 2007; Ho and Phillips, 2009). The use of multiple calibrations allows a better modelling of rate evolution across the branches in the tree, and should be preferably distributed at both deep and shallow divergences (Hedges and Kumar, 2009). Besides the number and distribution of calibrations, their quality is also important, both in the accuracy of their estimated age and their proximity to the true evolutionary divergence being calibrated (Hedges and Kumar, 2004). The Bayesian uncorrelated relaxed clock method implements a sophisticated approach to incorporate the information from such calibrations as prior distributions for selected internal nodes, including different distributions and with either hard or soft bounds (Drummond et al., 2006).

The software Beast v.1.6.1 (Drummond and Rambaut, 2007), which implements a uncorrelated relaxed clock method, was used to estimate divergence times among major frog lineages based on molecular data. The combined reduced dataset of study III was used, as in the phylogenetic

analyses (Table 3.1), and the tree topology was constrained to the best ML tree (Fig. 4.10) by removing the operators that act on tree topology from the xml file. The Yule process was used to describe cladogenesis, and independent GTR+I+Γ models were applied for each of the five data partitions. The final Markov chain was run twice for 100 million generations, sampling every 10,000 generations and the first million was discarded as part of the burnin process, according to the convergence of chains checked with Tracer v. 1.5. (Rambaut and Drummond, 2009). The effective sample size of all the parameters was above 200 (Drummond and Rambaut, 2007).

Seven calibration points were used as priors for divergence times of certain splits, using a lognormal distribution of prior probability (see Fig. 4.12), as it is regarded the most appropriate for modelling paleontological information (Hedges and Kumar, 2004; Drummond et al., 2006). Calibration points were chosen based on the online resource Lisanfos KMS v. 1.2 (Martín and Sanchíz, 2010). Fossils provided hard minimum bounds (offset) and mean and standard deviations (SD) were chosen so that the 95% credibility interval (CI) limit corresponds to a soft maximum bound.

A) Sauropsida–Synapsida split: Offset=312.3 mya as the minimum age for *Hylonomus* (Benton and Donoghue, 2007); log mean=2.2; log SD=0.424.

B) Archosauromorpha–Lepidosauromorpha split: Offset=259.7 mya as the minimum age for *Protorosaurus* (Benton and Donoghue, 2007); log mean=2.4468; log SD=0.756.

C) Cryptobranchidae–Hynobiidae split: Offset=145.5 mya. Following Roelants et al. (2007), a more conservative age estimate of *Chunerpeton* was used (Gao and Shubin, 2003); log mean=3.5; log SD=1.014. The soft maximum was 321 mya, corresponding to the upper 95% IC of this split in San Mauro (2010), as a more conservative upper bound than the one suggested by Marjanović and Laurin (2007).

D) Anura–Caudata split: Offset=249, which is the minimum age of *Triadobatrachus*, a stem Anura, (Rage and Roček, 1989); log mean=3.7; log SD=0.351; the soft maximum was 321 mya (see C).

E) Branching of Discoglossoidea: Offset=161.2 mya as the Middle-Late Jurassic boundary (Ogg et al., 2008) corresponding to *Eodiscoglossus*, the first known Discoglossoidea (Evans et al., 1990); log mean=3.6; log SD=0.532. The upper 95% IC value (soft maximum) was 249 mya (*Triadobatrachus*).

F) Branching of Pipoidea: Offset=145.5 mya as the minimum age for *Rhadinosteus*, the first-known pipoid (Henrici, 1998; Martín and Sanchíz, 2010); log mean=3.45; log SD=0.668. Soft maximum=249 mya (*Triadobatrachus*).

G) *Calyptocephalella–Lechriodus* split: Offset=52.8 mya as the oldest-known fossil of *Calyptocephalella* (Báez, 2000; Martín and Sanchíz, 2010); log mean=4.2; log SD=0.2. The age of *Rhadinosteus* (145.5 mya; see F) was used as a conservative soft maximum.

3.6. Methods for comparative biology

3.6.1. Alizarin-stained anatomical preparations

The classical technique of differential staining of bone and cartilage is typically used for studying the skeletal morphology of small vertebrates. First, the specimen must be fixed (using formaldehyde or ethanol) and cleared using a potassium hydroxide (KOH) solution, so that the soft tissues become transparent. Then, alizarin red S and alcian blue are used to stain differentially calcified tissues (in red) and cartilage (in blue) (Hildebrand, 1968).

Using the above technique, a comparative anatomical study of the larynx structures of several frog species was performed, in order to further understand the morphological basis of sound production in Pipidae. The anatomical study of larynges included several pipids (*Xenopus laevis, Hymenochirus boettgeri* and *Pseudhymenochirus merlini)* and *Bombina bombina* (representing non-pipid frogs for a comparative view). Specimens were sacrificed using an overdose of MS222, fixed in formalin, and differentially stained for bone and cartilage with alizarin red S and alcian blue, respectively, following a procedure suited for amphibians (Hanken and Wassersug, 1981).

3.6.2. Behavioural observations

Behavioural observations allowed the documentation of the particular sound produced by the pipid species *Pseudhymenochirus merlini*. Specimens of different species of the family Pipidae were obtained from the pet trade and kept at different times between 1985 and 2011. Observations of calling specimens of *P. merlini* were made on captive specimens in *ca.* 100 x 30 x 20 cm aquaria. All observations refer to specimens in breeding conditions, without external (hormone) stimulus. Video sequences of calling specimens were recorded in 2010 with a Sony DCR-SR30 camera. Spectral and temporal variables of the recorded sounds were analyzed using Cooledit 96 software (Syntrillium). Sonograms were constructed using the package seewave (Sueur et al., 2008) in the R environment (R Development Core Team, 2009).

3.6.3. Statistical analyses

In model-based phylogenetic methods, substitution rates are transformed into branch lengths assuming a given model of evolution, and therefore, testing differences in branch lengths is equivalent to comparing evolutionary rates among different taxa. As an additional approach to the relative-rate test, branch lengths of trees assuming a GTR+I+Γ model were compared between neobatrachian and non-neobatrachian frogs.

The separate mt and nuclear datasets from the third study (including all newly generated sequence data and a wide representation of the available data for frogs) were used, and both included all codon positions of protein-coding genes. Using RAxML v. 7.0.4 (Stamatakis, 2006), model parameters and branch lengths were separately optimized for the two datasets, constraining the topology to the preferred maximum likelihood tree of figure 4.10.

In order to compare the branch-specific bias of mt versus nuclear branch lengths, the ratio between branch lengths (mt/ nuclear) was calculated for each individual internal and terminal branch in the tree. The ratios between mt and nuclear branch lengths were subjected to a one-way ANOVA, after being log-transformed to meet the assumptions of normality and homogeneity of variance. Further analyses involved orthogonal contrasts between branches of non-neobatrachians, basal neobatrachians and derived neobatrachians (see 3.7.3 Relative rate tests for details of the species included in each of the three groups). All statistical analyses were performed with IBM SPSS Statistics, release 19.0.0.1.

3.7. Databases, trait evolution, and molecular evolution

3.7.1. The MitoZoa database

MitoZoa (http://mi.caspur.it/mitozoa; Lupi et al., 2010; D'Onorio de Meo et al., 2012) is a curated database of complete or nearly complete mt genomes specifically designed to assist comparative studies of architectural features of mt genomes. The data included in this dataset is reasonably reliable given that all the included mt genomes have been checked for missannotations that appear in some records in RefSeq (Pruitt et al., 2009; Hassanin et al., 2010). MitoZoa provides a suite of tools that allow a quick survey and straightforward study of gene order, non-coding regions and gene content (among other functionalities) between different taxa.

The newly sequenced frog mt genomes were compared against the MitoZoa database to inspect for similar gene orders among metazoans. Particularly, the MitoZoa provided comparative information on gene order in vertebrates (1,409 vertebrate mt genomes included as of February 2010), in order to assess the convergence with the gene order found in *Leiopelma archeyi*. GenBank (Benson et al., 2010) was also used to complement the above search. The MitoZoa database constantly adds new information and corrects detected errors, and thus, searches performed in 2010 and 2011 used releases 2.0 and 7.1, respectively.

3.7.2. BayesTraits

BayesTraits (by M. Pagel and A. Meade; http://www.evolution.rdg.ac.uk/BayesTraits.html) is a computer package for performing analyses of trait evolution among groups of species using a reference phylogeny. Here, the BayesMultistate (Pagel et al., 2004) program was used to model the evolution of the mechanism for sound production and the most likely ancestral states in the family Pipidae. The maximum likelihood approach was used to fit a continuous-time Markov model to the data, in which the discrete character can adopt two states (0, 1) and the instantaneous rates of evolution along branches were chosen to make the observed data most likely given the underlying phylogeny (Pagel, 1997). Both model parameters (rates of change between states) and ancestral states of characters at internal nodes were estimated by maximum likelihood, accounting for branch lengths in the phylogenetic tree (Pagel, 1997).

3.7.3. Relative-rate tests

The relative-rate test (Sarich and Wilson, 1973) evaluates the molecular clock hypothesis, comparing whether two or more groups of sequences of interest differ in their rates of evolution (Lemey and Posada, 2009). The program RRTree (Robinson-Rechavi and Huchon, 2000) extends the method of Li and Busquet (1992) and performs a relative rate test by comparing mean rates between lineages relative to a common outgroup, taking phylogenetic relationships into account by topological weighting (Robinson et al., 1998).

Three salamander species were used as outgroup, and frogs were divided into three groups: basal neobatrachian lineages (*Calyptocephalella, Heleophryne, Lechriodus* and *Sooglossus*), derived neobatrachians (Nobleobatrachia and Ranoides) and non-neobatrachian relatives (Amphicoela, Discoglossoidea, Pipoidea, and Pelobatoidea). Several relative-rate tests were performed in order to compare substitution rates of mt and nuclear genes among these three groups, as well as between all neobatrachians and non-neobatrachians, computing genetic distances with the Kimura 2 parameter model (Kimura, 1980), which allows for different transitions and transversion rates per site. Nucleotide alignments of single genes (with all three codon positions for protein-coding genes) were used, as well as the combination of all mt and all nuclear genes. The same tests were repeated with amino acid data, using an adaptation of the Jukes-Cantor method (Robinson-Rechavi and Huchon, 2000).

3.7.4. Estimating selection on alignments of protein-coding sequences

In protein-coding genes, nucleotide substitutions can be divided into synonymous and non-synonymous, depending respectively, on whether the encoded amino acid is altered or not. Non-synonymous substitutions can affect the protein function and they are more likely to influence the fitness of the individual, while synonymous substitutions do not change the encoded amino acid. A common method to study selective pressures on aligned protein-coding genes is by measuring the relative rate between the number of synonymous (silent) substitutions per synonymous (dS) site and the number of non-synonymous (replacement) substitutions per non-synonymous site (dN), also known as the selective coefficient (dN/ dS or ω) (Nielsen and Yang, 1998; Pybus and Shapiro, 2009). Under negative (or purifying) selection, non-synonymous changes accumulate more slowly than synonymous changes due to their lower fitness ($\omega < 1$); while under positive selection (diversifying or directional) the converse is true ($\omega > 1$) (Li et al., 1985). Both silent and replacement substitutions would accumulate at the same rate under neutral evolution ($\omega \sim 0$) (Li et al., 1985).

Using probabilistic codon substitution models in a maximum likelihood framework, ω coefficients can be estimated in a number of ways for a set of aligned sequences, in order to detect selective pressures on specific sites, lineages or sites within particular lineages (Kosakovsky Pond et al., 2009). Branch models (Yang, 1998a; Yang and Nielsen, 1998) that allow the ω ratio to vary among different branches in the phylogeny are useful to detect changes in selection acting on particular lineages. To examine whether the lineages of interest have a different ω coefficient compared to the remaining lineages, two models can be constructed: a null model assuming the same ω along all branches in the phylogeny and an alternative model in which the branches of interest are allowed to have a different ω. Then, the statistical significance of the difference can be calculated by means of a LRT, given that the two models are nested (Anisimova et al., 2001).

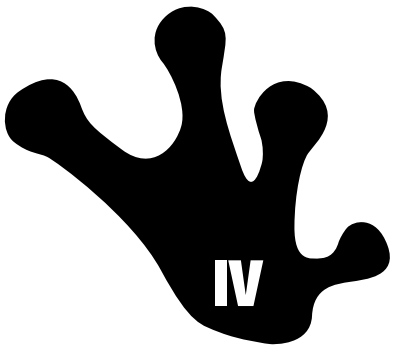
The branch models (Yang, 1998a; Yang and Nielsen, 1998) were used to investigate whether the acceleration of substitution rates in neobatrachians was due to changes in selection in this particular lineage. PAML v. 3.15 (Yang, 1997) was used to estimate ω values and likelihoods of all different models based on the preferred topology (Fig. 4.10) and sequence information both from single gene alignments and from the combination of all mt or all nuclear genes. Branch lengths were first optimized for each dataset under the null model and were fixed when all other parameters were estimated under alternative models.

The null model had a single ω value for all branches, and it was compared against four alternatives, which allowed a second ω value on (i) the stem branch of Neobatrachia, (ii) all neobatrachian branches, (iii) Ranoides, or (iv) Nobleobatrachia (all including their stem branches). Given that branch models assume that all background branches share the same ω, and in order to gain insight into the obtained results, we additionally allowed ω to vary on main non-neobatrachian lineages (including their stem branch): (v) Amphicoela, (vi) Discoglossoidea, (vii) Pipoidea, (viii) Pelobatoidea, or (ix) the stem branch of Pelobatoidea. These later additional models were compared against the null model by likelihood ratio tests, and, in addition, all 10 (non-nested) models were compared simultaneously using the AIC (Akaike, 1973).

3.7.5. Functional analysis of neobatrachian amino acid synapomorphies

We further examined whether the higher substitution rates and/ or distinct selective regimes of neobatrachians could be associated or caused by a disproportionally higher number of synapomorphic amino acid changes in this lineage. The program MrFunction (Abascal et al., details will be published elsewhere) was used to identify neobatrachian-specific molecular synapomorphies. Briefly, based on single-gene protein alignments, ancestral character states were reconstructed with MrBayes v. 3.1.2 (Ronquist and Huelsenbeck, 2003) for the last common ancestors of (i) Neobatrachia, (ii) its closer sister-group (Pelobatoidea), and (iii) the last common ancestor of both. Then, the three hypothetical ancestral sequences were compared to identify synapomorphic amino acid changes in both clades, taking only into account the sites with reliably reconstructed states (it was empirically found that $p > 0.75$ offered a good balance between the number of predictions and their corresponding reliability). A two-sided binomial test was used to assess whether the observed distribution of synapomorphies between Neobatrachia and Pelobatoidea was significantly different from that expected under the null hypothesis (*i.e.,* that the number of synapomorphies is the same in the last common ancestor of both clades).

In addition, to further understand if molecular synapomorphies of neobatrachians (or pelobatoideans) were associated to particular regions of the proteins, we predicted the accessibility to solvent and the occurrence at the different trans-membrane regions for each of the identified sites. Solvent accessibility was calculated through BLAST searches against the PDBFINDER2 database (Hooft et al., 1996), and trans-membrane helices of genes were predicted using TMHMM v.2.0 (Krogh et al., 2011).



RESULTS

4.1. The complete mitochondrial genome of the relict frog *Leiopelma archeyi*: insights into the root of the frog tree of life

4.1.1. Mitochondrial genome organization and structural features

The complete nucleotide sequence of the light strand of the mt genome of *L. archeyi* was determined, with a total length of 16,593 bp. The gene content was similar to most metazoan mt genomes (Boore, 2000); all tRNA genes could be folded into typical cloverleaf secondary structures with the known exception of *trnS–(AGY);* and the putative origin of replication of the light strand (between the *trnN* and *trnC* genes) had the potential to fold into a stem-loop secondary structure. Most protein-coding genes started with the codon ATG, with the exception of *cox1* (GTG), *nad1* (CTC), and *nad6* (TTG). Some genes had complete stop codons (TAA in *atp6* and *atp8, nad4L,* and *nad5*; TAG in *cob, nad1,* and *nad2*; AGG in *nad6*), whereas others (*cox1, cox2, cox3, nad3,* and *nad4*) ended with an incomplete codon (a single T).

The gene order found in *L. archeyi* departs from the consensus order of vertebrates (Boore, 1999; Gissi et al., 2008): *cob* and *trnT* genes are located immediately downstream of *nad5* gene; and *nad6, trnE,* and *trnP* genes are located between the control region and the *trnF* gene (Fig. 4.1). No changes in coding strand were observed for the rearranged genes. The control region is 858 bp-long, and it contains three conserved sequence blocks (CSB-1, CSB-2, CSB-3; Walberg and Clayton, 1981) that participate in the formation of a proper RNA primer in the process of replication of the mt DNA (Fernández-Silva et al., 2003), as well as three putative termination-associated

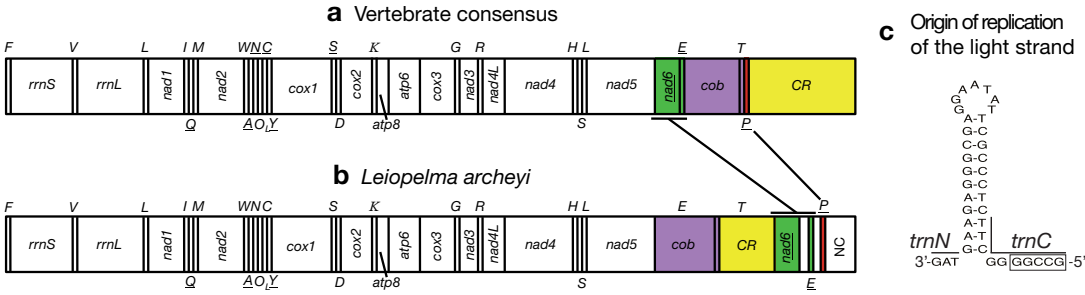


Fig. 4.1. Gene organization of the consensus mt gene order of (a) vertebrates, and (b) *L. archeyi.* Underlined genes are encoded by the light strand. Rearranged genes are coloured and lines indicate translocation of genes. (c) Proposed secondary structure for the origin of replication of the light strand of *L. archeyi.* The 5'-GCCGG-3' motif is indicated by a box and the lines show the overlapping regions with flanking tRNA genes.

sequences (TAS-1, TAS-2, TAS-3; Doda et al., 1981; MacKay et al., 1986). A 219 bp-long non-coding region was found between the *trnP* and *trnF* genes (Fig. 4.1). This region has two 46 bp-long non-tandem repeats separated by 27 nucleotides, 9 of which are also displayed by the two non-tandem repeats. Other intergenic spacers occurred between the *nad6* and *trnE* genes (15 bp) and between the *trnE* and *trnP* genes (17 bp). The putative origin of replication for the light strand was located within the *WANCY* tRNA cluster, between the *trnN* and *trnC* genes, and had the potential to fold into a stem-and-loop secondary structure (Fig. 4.1). The 5'-GCCGG-3' motif, which in human mt DNA replication is involved in the transition from RNA to DNA synthesis (Hixson et al., 1986), is entirely conserved in *L. archeyi.*

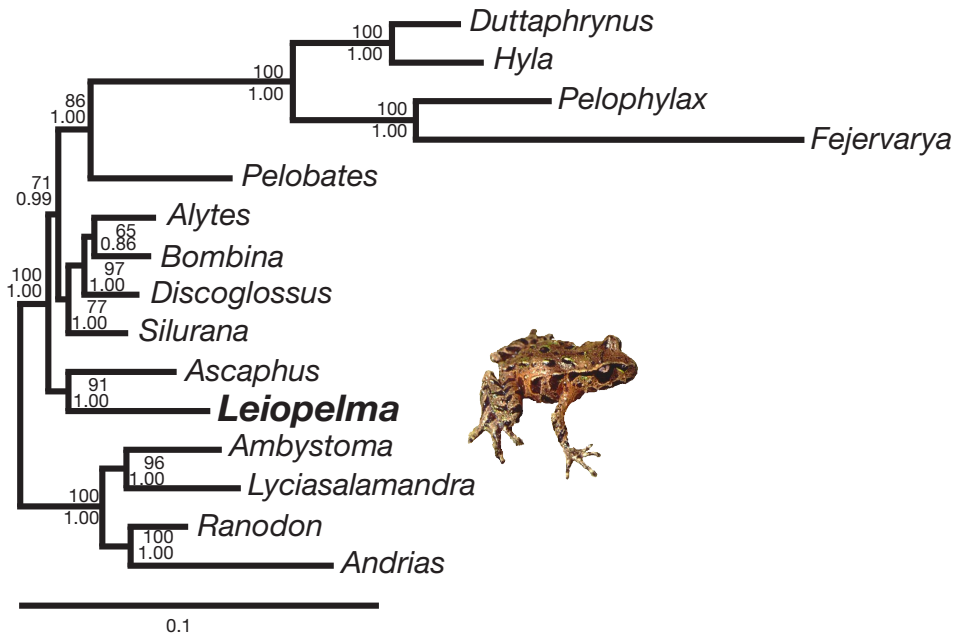


Fig. 4.2. Anuran phylogeny (maximum likelihood phylogram) inferred from a single concatenated data set of the deduced amino acid sequences of all mt protein-coding genes encoded by the heavy strand. The numbers above each node represent support for Bayesian inference (posterior probabilities; upper value) and maximum likelihood (1000 replicates bootstrap proportions, in percent; lower value). Scale bar represents substitutions · site^{-1}. Picture by D. M. Green, showing a specimen of *L. archeyi* of the Whareorino Forest, west of Te Kuiti, New Zealand.

4.1.2. Phylogenetic analyses

Sites of ambiguous alignment as well as those with the fastest substitution rates were excluded rendering a final matrix of 2,498 amino acid positions. Of these, 1,703 were invariant and 441 were parsimony-informative. Both maximum likelihood (-*ln*L = 15,839.46) and Bayesian inference (-*ln*L = 15,823.80 for run 1; -*ln*L = 15,823.93 for run 2) methods arrived at the same tree topology and differences were only observed in branch lengths and levels of support (Fig. 4.2).

The reconstructed tree strongly supported the sister group relationship between *Leiopelma* and *Ascaphus*, and the basal position of this clade as sister to all other anuran lineages (Fig. 4.2). Additionally, non-neobatrachian frogs were found to be paraphyletic with respect to Neobatrachia, thus rejecting the "Archaeobatrachia" hypothesis. Pipoidea (represented by *Silurana*) was recovered as the sister group of Discoglossoidea, and both together as the sister group of a clade composed of Pelobatoidea + Neobatrachia, both relationships receiving strong support (Fig. 4.2). Internal relationships within Discoglossoidea were poorly resolved (*Discoglossus* was recovered as sister to *Alytes* + *Bombina*, but with low statistical support; Fig. 4.2). Within Neobatrachia, two recognized clades (Hyloides and Ranoides, *sensu* Frost et al., 2006) were recovered with high support (Fig. 4.2).

According to the AU tests for seven alternative rooting and branching phylogenetic hypotheses (Table 4.1.), the mt sequence data rejected any topology that involved changes in the position of the root, *i.e,.* hypotheses placing *Leiopelma* (p = 0.023), *Ascaphus* (p = 0.037), or *Silurana* (p = 0.017) alone at the base of the anuran tree. The "Mesobatrachia" hypothesis, which implies a sister group relationship of *Silurana* and *Pelobates* was also significantly rejected (p = 0.009). Alternative hypotheses placing Pipoidea as the second (San Mauro et al., 2005; Frost et al., 2006) or third (Roelants and Bossuyt, 2005; Roelants et al., 2007) major lineage branching off the anuran tree could not be rejected (Table 4.1). The recovered internal phylogenetic relationships of the discoglossoids represented in our tree were not significantly different from those recovered by San Mauro et al. (2004a), *i.e.,* a sister group relationship between *Alytes* and *Discoglossus* to the exclusion of *Bombina*; which is also supported by many other posterior studies (Roelants and Bossuyt, 2005; San Mauro et al., 2005; Frost et al., 2006; Gissi et al., 2006; Roelants et al., 2007

Table 4.1. Log-likelihood and p values from the approximately unbiased (AU) tests for seven alternative rooting and branching scenarios. References are given below each alternative hypotheses. Probability values below 0.05 (bold italics) indicate that the combined dataset allowed the rejection of the particular alternative hypothesis.

Alternative hypotheses	**-*ln*L**	**p value**
Unconstrained tree	15,865.149	0.813
Leiopelma basal to all other frogs	15,875.565	***0.023***
Ascaphus basal to all other frogs (Ford and Cannatella, 1993)	15,875.080	***0.037***
(Bombina + (Discoglossus + Alytes)) (San Mauro et al., 2004a; Roelants and Bossuyt, 2005)	15,867.385	0.517
(Pipoidea + (Discoglossoidea + (Pelobatoidea + Neobatrachia))) (Haas, 2003; San Mauro et al., 2005; Frost et al., 2006)	15,874.319	0.076
(Discoglossoidea + (Pipoidea + (Pelobatoidea + Neobatrachia))) (Roelants and Bossuyt, 2005; Roelants et al., 2007)	15,872.534	0.279
Silurana sister of *Pelobates* ("Mesobatrachia") (Ford and Cannatella, 1993)	15,888.376	***0.009***
Pipoidea (*Silurana*) basal to all other frogs (Maglia et al., 2001; Púgener et al., 2003)	15,884.661	***0.017***

4.1.3. Convergent rearrangements in the mitochondrial genome

The gene order found in *L. archeyi* is new in anurans. However, this gene arrangement is convergent with that found in two species of plethodontid salamanders (*Aneides flavipunctatus* and *Stereochilus marginatus*) and 14 species of eels (*Ophisurus macrorhynchos, Myrichthys maculosus, Coloconger cadenati, Derichthys serpentinus, Nessorhamphus ingolfianus, Cynoponticus ferox, Muraenesox bagio, Paraconger notialis, Ariosoma shiroanago, Conger myriaster, Nettastoma parviceps, Hoplunnis punctata, Facciolella oxyrhyncha,* and *Leptocephalus* sp.) among all available vertebrate mt genomes (Table 4.2) (Benson et al., 2010; Lupi et al., 2010).

Given the large number of possible gene rearrangements in the mt genome, gene order was thought to present low levels of homoplasy, thus making it a reliable character to infer phylogenetic relationships (Boore and Brown, 1998). However, this premise is challenged by the convergences in mt gene order observed between *L. archeyi*, pelothodontid salamanders and several species of eels. Moreover, it has been shown that convergent gene orders are more likely to occur in hot spots for gene order change in the mt genome (San Mauro et al., 2006). To test whether the region upstream the control region (hereafter, CR 5' region) represents another hot spot for gene order change, it is necessary to (i) examine the extent of convergence of mt gene order in this region among vertebrates, and (ii) assess whether convergent gene orders appear at a higher frequency than that expected under pure chance.

The different gene orders found in the CR 5' region of vertebrates involve five rearranged genes (*cob, trnT, nad6, trnE, trnP*), and ignoring changes in the coding strand, there are 120 (5!) possible orders for those five genes; so, in principle there is a large character space and low probability of convergence (Boore and Brown, 1998). However, if we assume that the observed gene orders are produced by the tandem duplication—random loss model (thought to be the main mechanism in vertebrate mitochondria; Boore, 2000), 32 (2^5) possible random selections from each duplicated gene are possible, and thus, less than a quarter of the theoretically possible rearrangements would occur according to this model (San Mauro et al., 2006). Six of these random selections restore the ancestral vertebrate consensus, and thus 27 different gene orders can be produced by a single event of tandem duplication—random loss for these five genes under study (Table 4.2). In addition, this means that approximately one-fifth of all the rearrangements are expected to be undetectable.

Table 4.2. The 32 possible rearrangements for the CR 5' region that can be explained by duplication of the ancestral order [*nad6*$_1$ *trnE*$_1$ *cob*$_1$ *trnT*$_1$ *trnP*$_1$] into [*nad6*$_1$ *trnE*$_1$ *cob1 trnT*$_1$ *trnP*$_1$ *nad6*$_2$ *trnE*$_2$ *cob*$_2$ *trnT*$_2$ *trnP*$_2$] and subsequent deletion of redundant gene copies. Among the available mt genome data, species with gene orders departing from the vertebrate consensus are shown (along with the bibliographical reference or RefSeq accession number). Gene orders that cannot be explained by a single transposition from the vertebrate consensus order are marked with "a".

Possible rearrangements	Present in
nad6$_1$ *E*$_1$ *cob*$_1$ *T*$_1$ *P*$_1$	Vertebrate consensus
nad6$_1$ *E*$_1$ *cob*$_1$ *T*$_1$ *P*$_2$	Vertebrate consensus
nad6$_1$ *E*$_1$ *cob*$_1$ *P*$_1$ *T*$_2$	*Rudarius ercodes* (fish; Yamanoue et al., 2009) *Dallia pectoralis* (fish; Ishiguro et al., 2003
nad6$_1$ *E*$_1$ *T*$_1$ *P*$_1$ *cob*$_2$	
nad6$_1$ *cob*$_1$ *T*$_1$ *P*$_1$ *E*$_2$	*Ventrifossa garmani* (fish; Satoh et al., 2006)
E$_1$ *cob*$_1$ *T*$_1$ *P*$_1$ *nad6*$_2$	
nad6$_1$ *E*$_1$ *cob*$_1$ *T*$_2$ *P*$_2$	Vertebrate consensus
nad6$_1$ *E*$_1$ *T*$_1$ *cob*$_2$ *P*$_2$	
nad6$_1$ *E*$_1$ *P*$_1$ *cob*$_2$ *T*$_2$	
nad6$_1$ *cob*$_1$ *T*$_1$ *E*$_2$ *P*$_2$	
nad6$_1$ *cob*$_1$ *P*$_1$ *E*$_2$ *T*$_2$ *a*	
nad6$_1$ *T*$_1$ *P*$_1$ *E*$_2$ *cob*$_2$	
E$_1$ *cob*$_1$ *T*$_1$ *nad6*$_2$ *P*$_2$	
E$_1$ *cob*$_1$ *P*$_1$ *nad6*$_2$ *T*$_2$ *a*	
E$_1$ *T*$_1$ *P*$_1$ *nad6*$_2$ *cob*$_2$ *a*	
cob$_1$ *T*$_1$ *P*$_1$ *nad6*$_2$ *E*$_2$	Aves (100 species from different families; e.g., Mindell et al., 1998) *Rhineura floridiana* (reptile; Macey et al., 2004)
nad6$_1$ *E*$_1$ *cob*$_2$ *T*$_2$ *P*$_2$	Vertebrate consensus
nad6$_1$ *cob*$_1$ *E*$_2$ *T*$_2$ *P*$_2$	
nad6$_1$ *T*$_1$ *E*$_2$ *cob*$_2$ *P*$_2$	*Lampetra fluviatilis* (lamprey; NC_001131) *Petromyzon marinus* (lamprey; Lee and Kocher, 1995
nad6$_1$ *P*$_1$ *E*$_2$ *cob*$_2$ *T*$_2$	
E$_1$ *cob*$_1$ *nad6*$_2$ *T*$_2$ *P*$_2$	
E$_1$ *T*$_1$ *nad6*$_2$ *cob*$_2$ *P*$_2$ *a*	
E$_1$ *P*$_1$ *nad6*$_2$ *cob*$_2$ *T*$_2$ *a*	
cob$_1$ *T*$_1$ *nad6*$_2$ *E*$_2$ *P*$_2$	*Leiopelma archeyi* (frog; this study) *Ophisurus macrorhynchos* (conger eel; Inoue et al., 2004) *Myrichthys maculosus* (conger eel, NC_013635) *Coloconger cadenati* (conger eel, NC_013606) *Derichthys serpentinus* (conger eel, NC_013611) *Nessorhamphus ingolfianus* (conger eel, NC_013608) *Cynoponticus ferox* (conger eel, NC_0136179) *Muraenesox bagio* (conger eel, NC_013614) *Paraconger notialis* (conger eel; NC_013630) *Ariosoma shiroanago* (conger eel, NC_013632) *Conger myriaster* (conger eel; Inoue et al., 2001) *Nettastoma parviceps* (conger eel, NC_013625) *Hoplunnis punctata* (conger eel, NC_013623) *Facciolella oxyrhyncha* (conger eel, NC_013621) *Leptocephalus* sp. (conger eel, NC_01615) *Stereochilus marginatus* (salamander; Mueller et al., 2004) *Aneides flavipunctatus* (salamander; Mueller et al., 2004)
cob$_1$ *P*$_1$ *nad6*$_2$ *E*$_2$ *T*$_2$ *a*	
T$_1$ *P*$_1$ *nad6*$_2$ *E*$_2$ *cob*$_2$	
nad6$_1$ *E*$_2$ *cob*$_2$ *T*$_2$ *P*$_2$	Vertebrate consensus
E$_1$ *nad6*$_2$ *cob*$_2$ *T*$_2$ *P2*	*Bipes tridactylus* (reptile; Macey et al., 2004) *Bipes biporus* (reptile; Macey et al., 2004) *Bipes canaliculatus* (reptile; Macey et al., 2004
cob$_1$ *nad6*$_2$ *E*$_2$ *T*$_2$ *P*$_2$	
T$_1$ *nad6*$_2$ *E*$_2$ *cob*$_2$ *P*$_2$	*Plethodon elongatus* (salamander; Mueller et al., 2004)
P$_1$ *nad6*$_2$ *E*$_2$ *cob*$_2$ *T*$_2$	Vertebrate consensus
nad6$_2$ *E*$_2$ *cob*$_2$ *T*$_2$ *P*$_2$	

The different mt gene orders involving the CR 5' region (shown in Table 4.2) were identified through searches in the MitoZoa (Lupi et al., 2010) and GenBank (Benson et al., 2010) databases. From the 1,409 vertebrate mt genomes available in MitoZoa, the vast majority (1,100 entries) conformed to the vertebrate consensus. The remaining entries presented a distinct gene order, and 116 out of these involved the CR 5' region. Among these 116 rearrangements, seven different types of derived gene orders were found (Table 4.2). If we ignore those taxa in which some genes were deleted or duplicated, these seven types of rearrangements can be explained by a single tandem duplication event.

Taking phylogeny into account, these seven types of derived gene orders are the result of at least 12 independent rearrangement events that took place during the evolution of vertebrates (Table 4.2). The identical arrangement found in the two fish genera *Dallia* (1st putative event of independent rearrangement) and *Rudarius* (2nd) seems to be a convergence due to the strong evidence of the distinct origin and monophyly of Esociformes (Nelson, 2006) and Acanthomorpha (Stiassny, 1986; Johnson and Patterson, 1993; Wiley et al., 2000; Miya et al., 2003), respectively. *Ventrifossa garmani* (Satoh et al., 2006) represents a distinct derived gene order (3rd) from all other vertebrates. Birds show different convergences in gene order, but a single independent event of rearrangement (4th) seems to have taken place at the base of the clade (Mindell et al., 1998). Nevertheless, this ancestral rearrangement of the mt genome of birds is convergent with that of the reptile *Rhineura floridana* (5th) (Macey et al., 2004). All lampreys reported to date (6th) share another divergent mt gene order. The frog *L. archeyi* (7th) has a convergent gene order with 14 species of eels and two species of plethodontid salamanders. Eels of the suborder Congroidei include species with mt gene orders both conforming to the consensus of vertebrates and with a similar gene order to *L. archeyi*, but for simplicity reasons, and because phylogenetic relationships within Congroidei still need to be resolved (Nelson, 2006), the latter was considered as a single and independent event of mt genome rearrangement (8th). As pointed out by Mueller et al. (2004), the gene orders found in the plethodontid salamanders *Stereochilus marginatus* (9th) and *Aneides flavipunctatus* (10th) have an independent origin. Additional derived mt gene orders correspond to three species of reptiles of the genus *Bipes* (11th) (Macey et al., 2004) and to the salamander *Plethodon elongatus* (12th) (Mueller et al., 2004).

The present case of the CR 5' region is fairly similar to that reported for the *WANCY* region (San Mauro et al., 2006), in the sense that both are close to origins of replication and involve five rearranged genes. Following San Mauro et al. (2006), the conditional probability of at least one convergence given 12 independent rearrangements in the CR 5' region is 0.95.

4.2. Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs, family Pipidae

4.2.1. Mitochondrial genome organization and structural features

The complete sequence of the light strand of the mt genome of the following four pipoids was determined for the first time: *Hymenochirus boettgeri, Pipa carvalhoi, Pseudhymenochirus merlini,* and *Rhinophrynus dorsalis*. The mt genome of *Xenopus laevis* was the first ever determined for an amphibian (Roe et al., 1985) but contained numerous minor sequencing errors probably due to technical constraints at that time. We therefore sequenced anew the full mt genome of this model species as well. The gene content and order in all pipoid species conformed to the consensus of vertebrates (Boore, 1999; Lupi et al., 2010); all tRNA genes could be folded into the typical cloverleaf secondary structure with the known exception of *trnS–(AGY);* and the putative origin of replication of the light strand had the potential to fold into a stem-loop secondary structure, and was located between the *trnN* and *trnC* genes in all species. Three conserved sequence blocks (CSB-1, CSB-2, CSB-3) were identified in the 3' end of the mt control region in all pipoid species. Notably, our sequence of *Rhinophrynus* (a specimen from Tenexpa, Pacific coast of Mexico) differs from a previously determined sequence (GenBank accession number DQ283109; from the Caribbean coast of Texas, US; Frost et al., 2006) by a high uncorrected p divergence of 9.9%, suggesting the existence of a second unrecognized species in this monotypic genus and family.

4.2.2. Phylogenetic analyses

Both maximum likelihood and Bayesian inference methods of phylogenetic reconstruction recovered fully congruent tree topologies for mt, nuclear, and combined datasets, with differences only in branch lengths and levels of support (Fig. 4.3). Five major clades were recovered within Anura (Fig. 4.3): Amphicoela (*Leiopelma* + *Ascaphus*, which were used to root the tree), Discoglossoidea, Pipoidea, Pelobatoidea and Neobatrachia. Non-neobatrachian frogs were recovered as successively branching lineages, with Discoglossoidea branching off after Amphicoela, followed by Pipoidea and Pelobatoidea. These relationships received high support values in the analysis of mt genomes and nuclear genes, and maximum support in the combined analysis (Fig. 4.3). Alternative phylogenetic placements of the Pipoidea were significantly rejected by AU tests (Table 4.3).

Fig. 4.3. Phylogenetic relationships among frogs (maximum likelihood phylograms) based on the analysis of concatenated DNA sequences of (a) mt genomes, (b) nine nuclear genes, and (c) combined mt genomes and nuclear loci. Numbers at nodes are support values from maximum likelihood bootstrap (1000 replicates; in percent) and Bayesian inference posterior probabilities. Scale bar represents substitutions · site^{-1}.

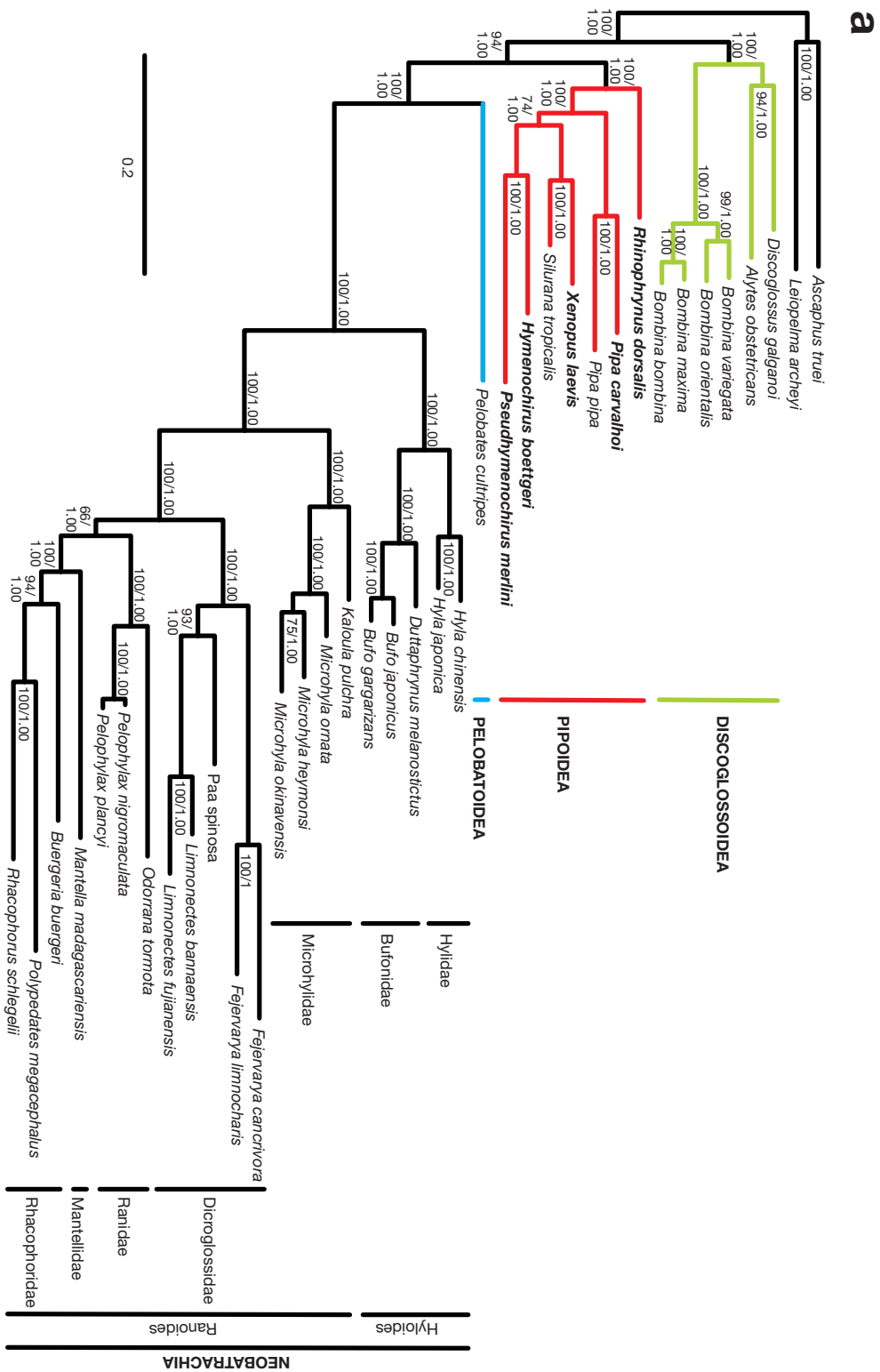
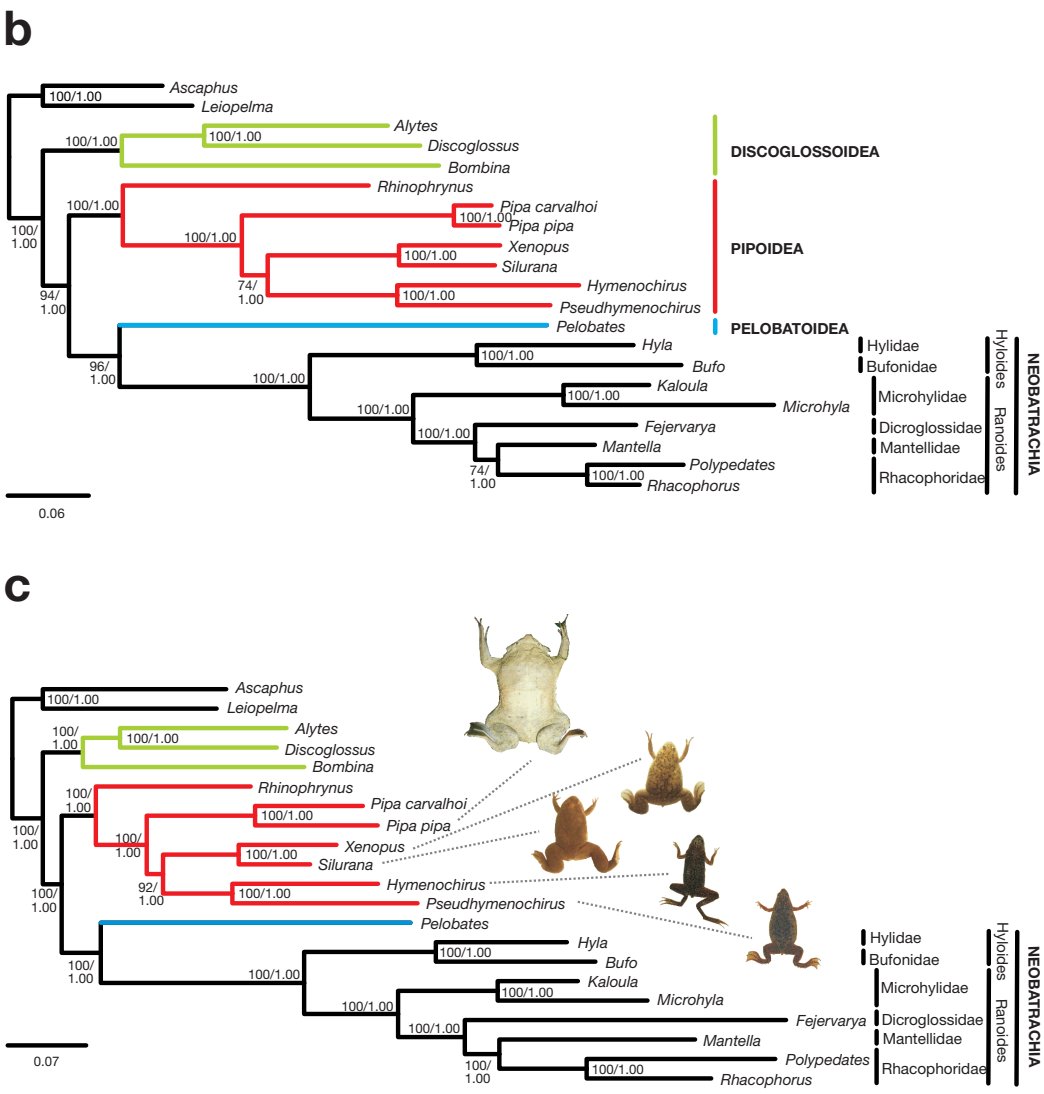


Fig. 4.3. Cont.



Within the Pipoidea, all data sets and phylogenetic analyses supported (i) *Rhinophrynus* (Rhinophrynidae) as the sister taxon of a monophyletic Pipidae, (ii) *Pipa* as sister group to all other extant pipid genera, and (iii) sister group relationships between *Xenopus* and *Silurana* (Dactylethrinae), and between *Hymenochirus* and *Pseudhymenochirus* (Hymenochirini) (Fig. 4.3). All alternative hypotheses could be significantly rejected, except that of a sister group relationship between *Pipa* and *Hymenochirus*, to the exclusion of *Xenopus* + *Silurana* (Table 4.3). In single gene analyses of nuclear data, *Pipa* was recovered as sister group to all other extant pipid genera by *bdnf, pomc, cxcr4, slc8a1* and *slc8a3* genes, whereas *rag1* and *rag2* genes recovered the *Xenopus* + *Silurana* clade and the *Hymenochirus* + *Pseudhymenochirus* clade in such basal position, respectively (Table 4.4).

Table 4.3. Results of the approximately unbiased (AU) tests using the combined matrix with all 37 mt and 9 nuclear genes. Bibliographical references of alternative hypotheses are given below each topology. Probability values below 0.05 (bold italics) indicate that the combined dataset allowed the rejection of the particular alternative hypothesis.

Alternative hypotheses	***-ln*L**	**p value**
Unconstrained tree	154,788.134	0.96
Phylogenetic position of Pipoidea within Anura		
(Pipoidea + (Discoglossoidea + (Pelobatoidea + Neobatrachia))) (Haas, 2003; San Mauro et al., 2005; Frost et al., 2006)	154,826.5114	***0.001***
"Mesobatrachia" hypothesis (Pipoidea + Pelobatoidea) (Ford and Cannatella, 1993; Hillis et al., 1993; García-París et al., 2003b)	154,836.3637	***2·10^{5}***
Pipoidea + Discoglossoidea (Gissi et al., 2006)	154,821.8280	***0.003***
Monophyly of Archaeobatrachia as in Hay et al. 1995 (Pelobatoidea+(Pipoidea+(Amphicoela+Discoglossoidea))+Neobatrachia)	154,869.7927	***4·10^{6}***
Internal relationships within Pipidae		
(*Xenopus* + (*Silurana* + (*Pipa* + (*Hymenochirus* + *Pseudhymenochirus*)))) (Cannatella and Trueb, 1988b)	155,352.7071	***1·10^{9}***
((*Pipa* + *Hymenochirus*) + (*Xenopus* + *Silurana*)) (Maglia et al., 2001; Púgener et al., 2003; Evans et al., 2004; Evans et al., 2005a; Trueb et al., 2005)	154,807.3520	0.089
Pseudhymenochirus basal in Pipidae	155,079.4269	***0.021***
(*Pseudhymenochirus + Hymenochirus*) basal in Pipidae	154,813.8851	***1·10^{6}***

Table 4.4. Summary of support for phylogenetic relationships among pipoids from single-gene ML bootstrap analyses of seven nuclear genes (data for two additional genes, rhodopsin and histone 3, is not shown because the analyzed fragments were very short). Strong support refers to proportions of non-parametric bootstrapping higher than 70%. Monophyly of Pipa refers to the monophyly of P. carvalhoi and P. pipa when the sequences of both species were available for the individual gene analyses, otherwise a hyphen is shown.

	bdnf	***cxcr4*, exon 2**	***rag1***
Monophyly of Pipoidea	not recovered	strong support	strong support
Monophyly of Pipidae	strong support	strong support	strong support
Internal relationships within Pipidae	*Pipa* is basal; weak support	*Pipa* is basal; strong support	Dactylethrinae is basal; weak support
Monophyly of *Pipa*	–	–	strong support
Dactylethrinae: *Xenopus + Silurana*	strong support	strong support	strong support
Hymenochirini: *Hymenochirus + Pseudhymenochirus*	strong support	strong support	strong support

	rag2	***slc8a1*, exon 2**	***slc8a3***	***pomc***
Monophyly of Pipoidea	strong support	strong support	strong support	not recovered
Monophyly of Pipidae	strong support	strong support	strong support	strong support
Internal relationships within Pipidae	Hymenochirini is basal; weak support	*Pipa* is basal; weak support	*Pipa* is basal; weak support	*Pipa* is basal; weak support
Monophyly of *Pipa*	–	strong support	–	–
Dactylethrinae: *Xenopus + Silurana*	strong support	strong support	strong support	strong support
Hymenochirini: *Hymenochirus + Pseudhymenochirus*	strong support	strong support	strong support	strong support

4.2.3. Evolution of the mechanism of sound production in Pseudhymenochirus *and other Pipidae*

In contrast to previous non-documented observations (Yager, 1996), we provide compelling behavioural data on *Pseudhymenochirus merlini* showing that individuals of this species, while calling, move a column of atmospheric air from the lungs through the glottis. We conclude that this movement of air almost certainly is causal for sound production in this species. A movie of a calling male of *Pseudhymenochirus merlini* can be found as an additional file in the published manuscript (http://www.biomedcentral.com/1471-2148/11/114/additional). Unlike all other extant pipid genera, all of which show a motionless calling, vocalizations in *Pseudhymenochirus* are clearly associated with intermittent constrictions of the posterior flanks and extension of the throat (Fig. 4.4). The observed sequence of movements further suggests that sounds are produced during expiration, *i.e.,* movement of the air from the lungs (Fig. 4.4).

Two different types of vocalizations were observed in *P. merlini*: male advertisement calls and release calls. Male advertisement calls were emitted underwater; whereas release calls were emitted when the observer gently clasped a male in the inguinal region. No female calls were heard, and no female release calls could be evoked when clasping unreceptive females, despite several attempts in different specimens.

Advertisement calls were emitted by submerged males sitting on the ground of the aquarium, in a posture with the head slightly turned upwards. During sound emissions, weak but very distinctly recognizable contractions of the flanks occurred (especially in the inguinal region), alternating with a slight inflation and deflation of the throat. One sequence started with the contraction of the flanks, and subsequently the throat became inflated. During this sequence, one note was emitted, clearly indicating an expiratory mechanism in which sound production relies on the air stream running from the lungs to the throat.

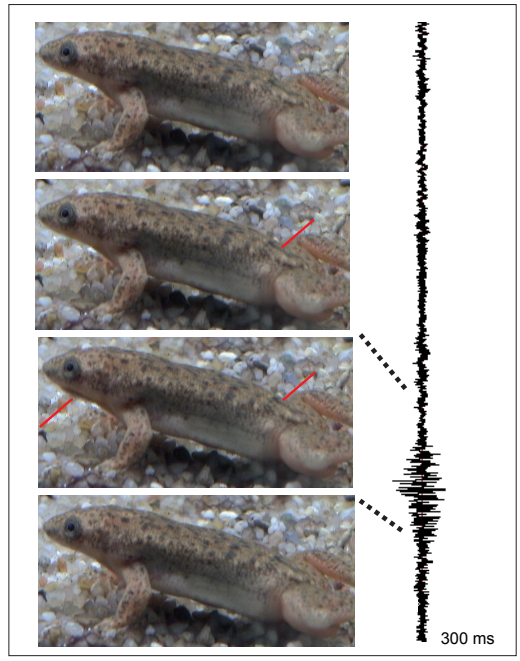


Fig. 4.4 Sound production in *Pseudhymenochirus merlini*. Time series of emission of one note in a male, showing movement of throat and flanks, indicative of movement of an air column.

The advertisement call is a rapid series of usually four, sometimes five, short non-melodious notes (Fig. 4.5a). The following description is based on recordings of a single male without hormonal stimulation, but other males were observed to emit similar calls. Call duration in four-note calls is 604-682 ms (mean 642 ± 22 ms; N=20); interval between calls 1930-3225 ms (2392 ± 363 ms; N=20). Note duration is 23-35 ms (29 ± 4 ms; N=20, measured on 7 different calls); duration of intervals between notes is 127-158 ms (138 ± 12 ms; N=20). No clear pulses can be recognized within each note. Frequency is 50-2200 Hz, dominant frequency about 690 Hz.

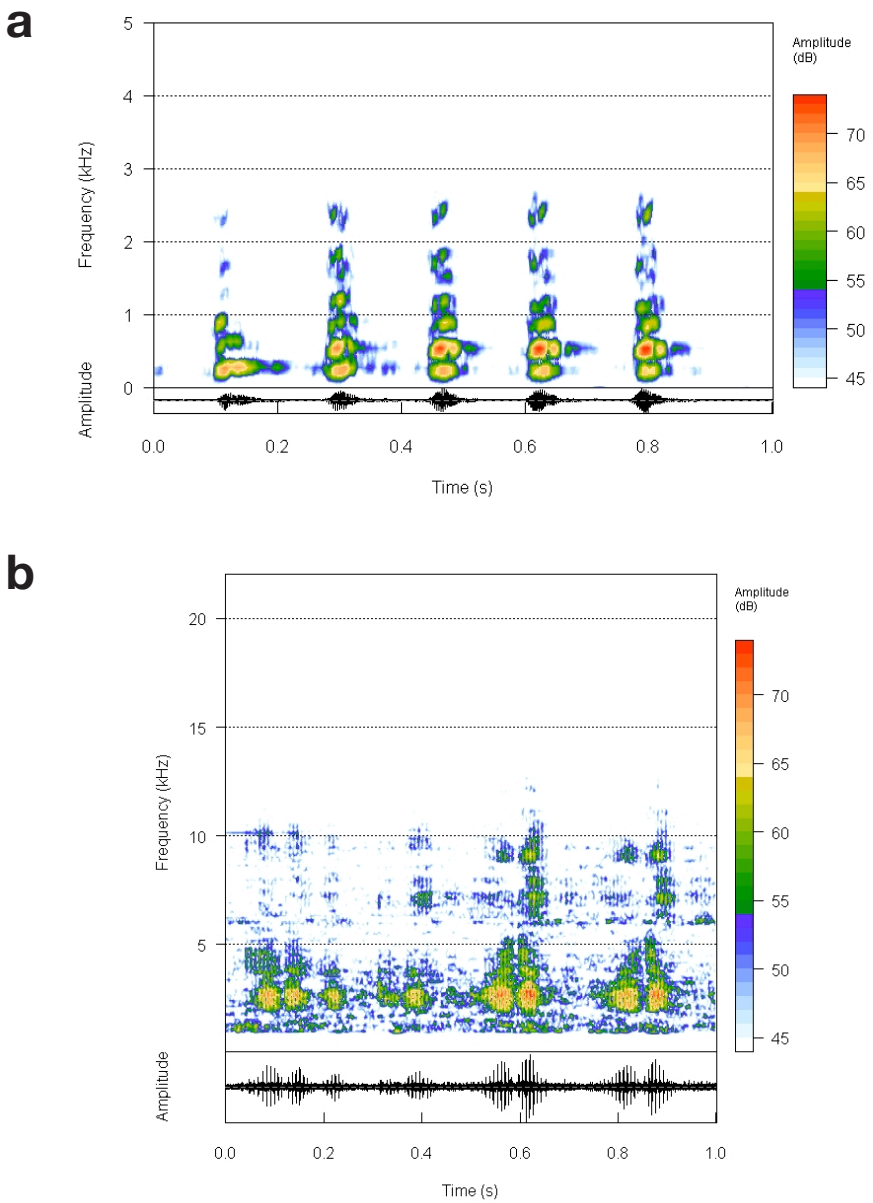


Fig. 4.5. Sonagram and oscillogram of (a) one advertisement call and (b) one release call (both with five notes) of *Pseudhymenochirus merlini*..

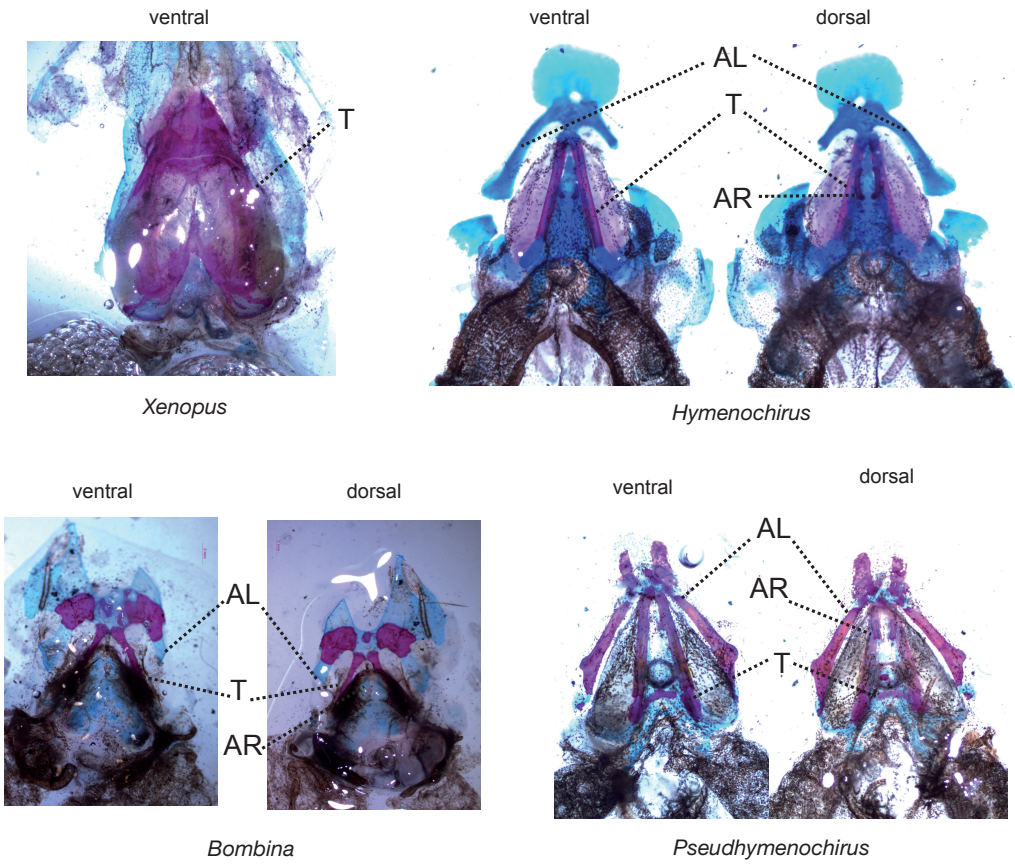


Fig. 4.6. Cleared and stained preparations using alizarin red–alcian blue, for the larynges of the pipid frogs *Pseudhymenochirus merlini, Hymenochirus boettgeri,* and *Xenopus laevis*, and the discoglossoid frog *Bombina orientalis.* Abbreviations: AL, alary processes of the hyoid plate; AR, arytenoid cartilages; T, thyrohyals (= posteromedial processes of hyoid). Colours denote calcified (red) vs. non-calcified cartilaginous (blue) structures. Note that in the three pipids, the whole box-like structure with numerous calcified elements (red stain) is the larynx (not marked), whereas the larynx of *Bombina* only consists of cartilaginous elements (blue stain) and soft tissue.

Males emitted release calls regularly when clasped. They are short series of rather irregular pulsed notes of variable duration (Fig. 4.5b). In one such release call, note duration was *ca.* 120-190 ms (exact limits between notes were difficult to define). Frequency was 1500-5000 Hz, with some bands also recognizable up to > 10000 Hz. Dominant frequency was 2550 Hz. Notes contained about 10-20 distinct pulses which often were arranged in two pulse groups. During the call, flank contractions were observed, suggesting that the sound is indeed produced by an airstream mechanism (Fig. 4.4). Pulse rate was about 140 per second. The call strongly reminded the advertisement calls of painted frogs of the genus *Discoglossus* which have two distinct pulse groups corresponding to an inspiratory and expiratory airstream (Weber, 1974; Glaw and Vences, 1991).

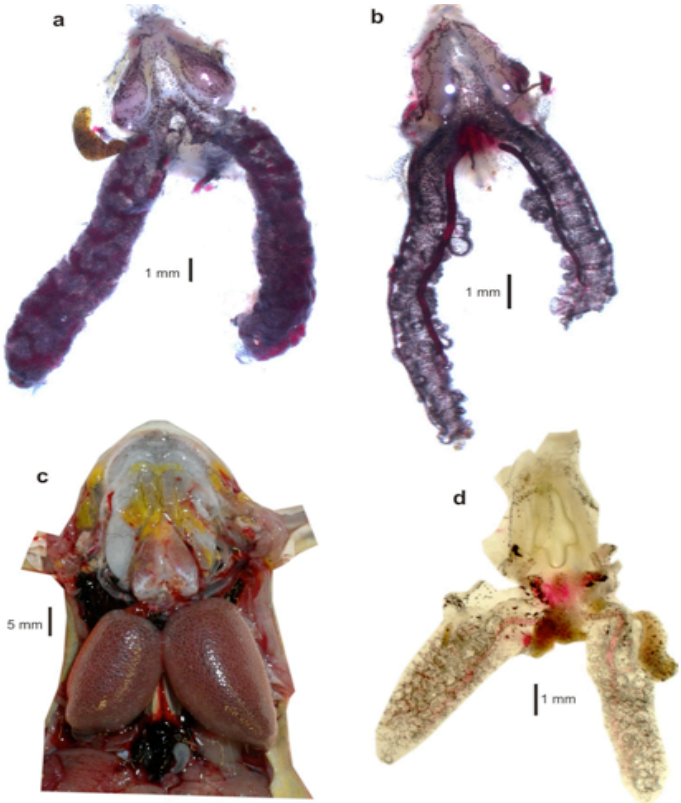


Fig. 4.7. Fresh preparations of larynx and lungs of (a) *Pseudhymenochirus merlini*, (b) *Hymenochirus boettgeri*, and (c) adult and (d) juvenile of *Xenopus laevis*. Note the similarity between *Pseudhymenochirus* and *Hymenochirus* in the elongate form of the lungs and superficially box-like larynx structure, as well as the greatly enlarged larynx of *Xenopus*.

Behavioural observations were complemented with morphological comparative analyses of the larynx structures of fresh anatomical (Fig. 4.7) and alizarin red–alcian blue stained preparations (Fig. 4.8) of different pipid genera and a discoglossoid. A detailed comparative anatomical and functional analysis of the larynx of *Pseudhymenochirus* is beyond the scope of the present study, but these observations illustrate several key points to further understand the call mechanism observed in this species. First of all, previous studies were confirmed, showing that the larynx in pipids is a prominent box-like structure surrounded by hard cartilage, which is (at least partially) ossified (Figs. 4.8 and 4.9) (Ridewood, 1897, 1900; Yager, 1992). In contrast, the larynx is neither enlarged nor ossified in other non-pipid frogs (*e.g., Bombina*, Fig. 4.6). In *Bombina* the thyrohyals are not directly connected to the larynx, while in the pipids it is an integral part of the box-like larynx structure. Previous anatomical descriptions of *Xenopus* (Ridewood, 1897; Yager, 1996) and *Hymenochirus* (Ridewood, 1900) were also confirmed. The larynx of *Xenopus*, is a highly ossified box made up by the thyrohyals, arytenoids, and cricoid cartilages, which are greatly expanded posteriorly. Furthermore, upon fresh anatomical dissections, the production of single clicks in the isolated larynx of *X. laevis* could be stimulated by gently touching and pressing the tendon muscles simultaneously on both sides of the larynx capsule, similar to what has been described for *X. borealis* (Yager, 1996).

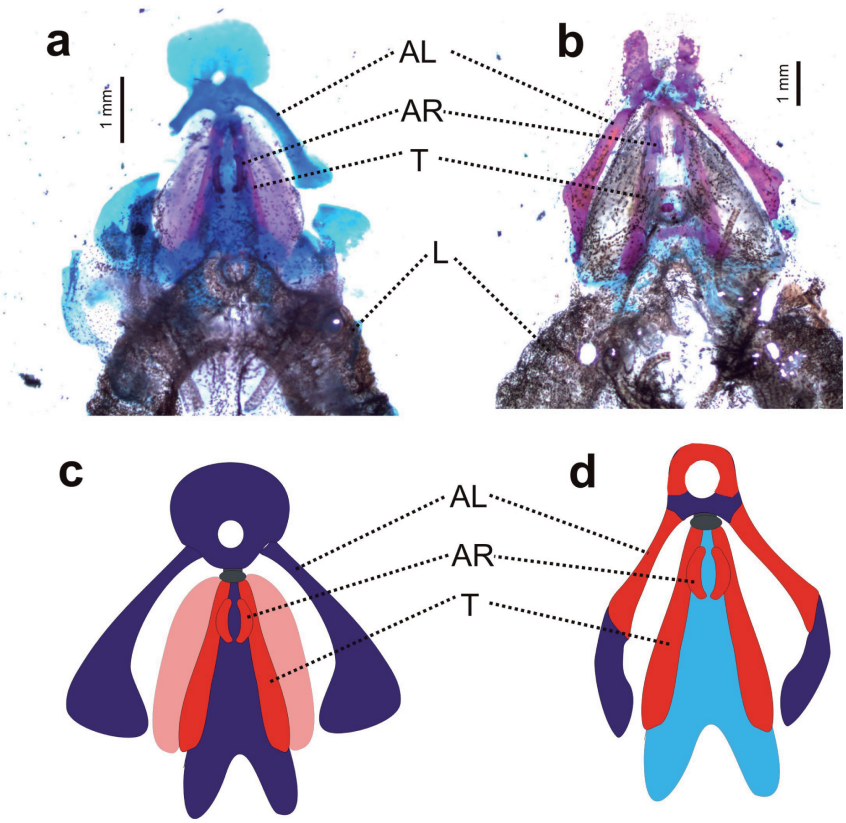


Fig. 4.8. Cleared and stained preparations of the larynx of (a) *Hymenochirus boettgeri* and (b) of *Pseudhymenochirus merlini* in dorsal view, showing a generally lower extension of cartilaginous and calcified structures surrounding the larynx in *Pseudhymenochirus*. Abbreviations: L, lungs; AL, alary processes of hyoid plate; AR, arytenoid cartilages; T, thyrohyals (= posteromedial processes of the hyoid plate). Schematic drawings represent main larynx structures in (c) *Hymenochirus* and (d) *Pseudhymenochirus*. Colours denote calcified (red) vs. non-calcified cartilaginous (blue) structures. Note the calcified alary process in *Pseudhymenochirus*. Drawings modified from Ridewood (1900) and Cannatella and Trueb (1988b).

In both *Hymenochirus* and *Pseudhymenochirus*, larynges show conspicuous and ossified thyrohyals that enclose the smaller arytenoid rods (Fig. 4.8). Both genera additionally share an elongate shape of lungs that reach the inguinal region and are tightly attached to the body wall (Fig. 4.7). However, there are two conspicuous differences between these two genera: (i) *Pseudhymenochirus* has ossified alary processes of hyoid plate, which form rods very similar to the thyrohyals (= posteromedial processes of the hyoid plate), and (ii) cartilage and calcified structures around the larynx are more extended and form an overall more compact laryngeal "box" structure in *Hymenochirus*, with calcified structures lateral to the thyrohyals and extensive cartilage visible in the glottis area (Fig. 4.8). Because all cleared-and-stained preparations were made from adult specimens that were sacrificed immediately previous to the clearing and staining procedure, it can be excluded that preservation artefacts could have caused these differences.

Overall, these morphological evidences clearly show the pipid nature of the larynx of *Pseudhymenochirus*. The overall pipid-like larynx in *Pseudhymenochirus* is clearly illustrated by the presence of the typical modified and ossified arytenoid cartilages and thyrohyals of pipids (Ridewood, 1897; Rabb, 1960). However, the larynx of *Pseudhymenochirus* appear much less robust than that of its sister genus *Hymenochirus* (Figs. 4.6 and 4.8), and thus we suggest that this fact would make the overall larynx more flexible and somehow permit a movement of air through it to produce vocalizations. However, whether vocal cords, which are absent in other pipids (Ridewood, 1897, 1900; Yager, 1992), are present in *Pseudhymenochirus*, or different structures are responsible for sound production during movement of the airstream awaits further detailed examination

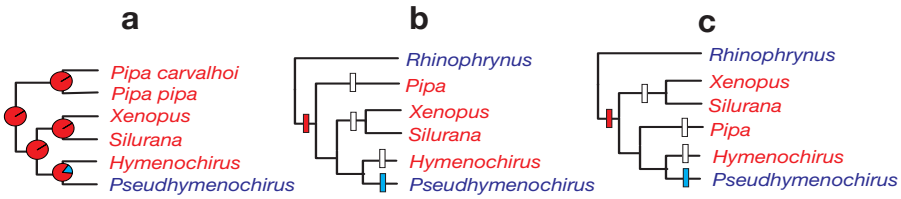


Fig. 4.9. (a) Reconstruction under maximum likelihood of ancestral character states of sound production mechanism (red without, and blue, with movement of air column) using BayesMultistate. (b) Preferred ancestral character state reconstruction of origin (red bar) and reversal (blue bar) of sound production mechanism; white bars represent the less parsimonious hypothesis of three independent origins of the implosion mechanism. (c) Same reconstruction under the alternative pipoid phylogeny suggested by morphology (*e.g.,* Púgener et al., 2003; Trueb et al., 2005).

Parsimony optimization of ancestral character states of sound production mechanism on the recovered hypothesis of pipid phylogenetic relationships, and on the only not significantly rejected alternative phylogenetic hypothesis (Fig. 4.9) supported homoplasy of the air-driven call in *Pseudhymenochirus*. Because pipid sound production appears to be linked to the adaptation to aquatic environment (Yager, 1996) and character states in fossil taxa are unknown, a single origin in the ancestor of Pipidae is assumed, followed by a subsequent reversal in *Pseudhymenochirus*, *i.e.,* two transformations. The alternative hypothesis would require assuming independent origin of the implosion mechanism in *Pipa*, *Xenopus* + *Silurana,* and *Hymenochirus*, and thus three transformations. BayesMultistate reconstructed the ancestral pipid character state as using the implosion mechanism (present in all pipids except *Pseudhymenochirus*), with a maximum likelihood probability > 0.999 (Fig. 4.9).

4.3. Phylogeny at the basis of modern frogs (Neobatrachia), and lineage-specific substitution rate heterogeneity of complete mitochondrial genomes and nine nuclear loci

4.3.1. Mitochondrial genome organization and structural features

The complete nucleotide sequence of the light strand of the mt genome of one pelobatoid (*Pelodytes punctatus*) and five neobatrachian species (*Heleophryne regis, Lechriodus melanopyga, Calyptocephalella gayi, Telmatobius bolivianus, Sooglossus thomasseti*), were determined anew, as well as the nearly complete mt genome of *Sooglossus sechellensis*. The gene content was identical to most metazoan mt genomes (Boore, 1999; Gissi et al., 2008). All tRNA genes could be folded into the typical cloverleaf secondary structure with the known exception of *trnS–(AGY).* The putative origin of replication of the light strand had the potential to fold into a stem-loop secondary structure and was located between the *trnN* and *trnC* genes in all species, with the exception of *H. regis* and *L. melanopyga* (see below). Three conserved sequence blocks (CSB-1, CSB-2, CSB-3) were identified in the 3' end of the control region of all species. The length of the control regions varied widely among the completely sequenced mt genomes (from 982 bp in *S. thomasseti* to > 3,866 bp in *T. bolivianus*), and included one or more tandem repeats. The actual length of the mt control regions of two species could not be accurately determined due to the presence of long tandem repeats (200-300 bp-long in *T. bolivianus* and 1,800-2,100 bp-long in *L. melanopyga*), and thus they were estimated from the length of the PCR product in the electrophoresis gel. The total length of the mt genomes of *T. bolivianus* and *L. melanopyga* was about 19500-19600bp and 21000-21300 bp, respectively.

The gene order of the mt genome in *Pelodytes punctatus* follows the consensus of vertebrates and other reported pelobatoids (Boore, 1999; Gissi et al., 2006). *Calyptocephalella gayi*, *Telmatobius bolivianus, Sooglossus thomasseti* and *S. sechellensis* conform to the consensus mt gene order of neobatrachians, which differs from the vertebrate consensus mt gene order in the translocation of *trnL–(CUN), trnT and trnP* genes from upstream the 5' end of the control region to form the *LTPF* tRNA cluster, which is typically located downstream the control region before the 5' end of the *rrnS* gene (Sumida et al., 2001). The mt genome of *Lechriodus melanopyga* follows the consensus neobatrachian gene order with the only exception of the location of the origin of replication of the light strand in a 218 bp-long intergenic spacer between *trnY* and *cox1* genes (Fig. 4.10). *Heleophryne regis* has a new gene order not reported in any other animal species

(Lupi et al., 2010), in which the *trnM* gene is pseudogenized (the anticodon has a deletion) in its ancestral location (*IQM* tRNA cluster) and the functional *trnM* appears within the *WANCY* cluster, which is rearranged as trn*A, trnN, trnC, trnM, trn, trnW, and trnY*, without changes in the coding strands (Fig. 4.10). The origin of replication of the light strand is located in a 165 bp-long intergenic spacer between *trnW* and *trnY* genes (Fig. 4.10). Notably, the *LTPF* cluster has a 385 bp-long intergenic spacer between *trnP* and *trnF* genes with no obvious sequence similarity (based on BLAST searches), and contains an 82 bp-long tandem repeat. Similarly to *Heleophryne regis,* other trnM pseudogenes have been reported in mt genomes of vertebrates: several members of the family Mantellidae (Kurabayashi et al., 2006; Kurabayashi et al., 2008), and some fishes (parrot fishes of the family Scaridae; Mabuchi et al., 2004, *Carapus bermudensis* [Ophidiiformes], Miya et al., 2003; and *Diaphus splendidus* [Myctophiformes]; Miya et al., 2001).

4.3.2. Phylogenetic analyses

The mt plus nuclear combined dataset consisted of 21,546 positions and 28 species, divided into 14 neobatrachians and 14 non-neobatrachians (not including the outgroups for timetree estimation; Table 3.1). Due to the long-branch attraction effect found in the mt dataset (see below), a combined reduced dataset of 11,136 bp was created, by retaining only more conserved positions (see Materials and Methods 3.4.2 and Table 3.1). Based on the combined reduced dataset with 5 partitions, both maximum likelihood (-*ln*L=76,155.99) and Bayesian inference methods (-*ln*L=76,547.00 for run 1; -*ln*L=76,548.29 for run 2) arrived to the same topology (Fig. 4.10). Five major clades of frogs were recovered with high support: non-neobatrachian lineages branched off successively as (i) Amphicoela, (ii) Discoglossoidea, (iii) Pipoidea and (iv) Pelobatoidea, which was the sister group of (v) Neobatrachia.

Leiopelma and *Ascaphus*, which were used to root the tree, were recovered as sister genera with high support; within Discoglossoidea, *Discoglossus* and *Alytes* were sister genera to the exclusion of *Bombina*, and Rhinophrynidae was the sister group of Pipidae, forming the clade Pipoidea. Within Pipidae, *Pipa* species branched off first, and *Hymenochirus* + *Pseudhymenochirus* were sister group to *Xenopus* + *Silurana*. *Pelodytes* was recovered as sister to *Pelobates*, forming the clade Pelobatoidea. Neobatrachia was confirmed as a monophyletic taxon, with *Heleophryne* as sister group to all other neobatrachians, although this relationship received support only in the Bayesian inference analysis (Fig. 4.10). The remaining neobatrachians are organized into two main clades. In one clade, *Lechriodus* and *Calyptocephalella* were recovered as sister taxa, and both as sister to Nobleobatrachia (Fig. 4.10). Within Nobleobatrachia, *Duttaphrynus* and *Telmatobius* were sister taxa to the exclusion of *Hyla* (Fig. 4.10). In the other clade, the two *Sooglossus* species included were grouped together (hence corroborating the genus *Sooglossus*) at the base of Ranoides (Fig. 4.10), even though support for this latter split was only moderate (maximal for Bayesian inference, but low for maximum likelihood). Ranoidean intra-relationships received high support (Fig. 4.10):

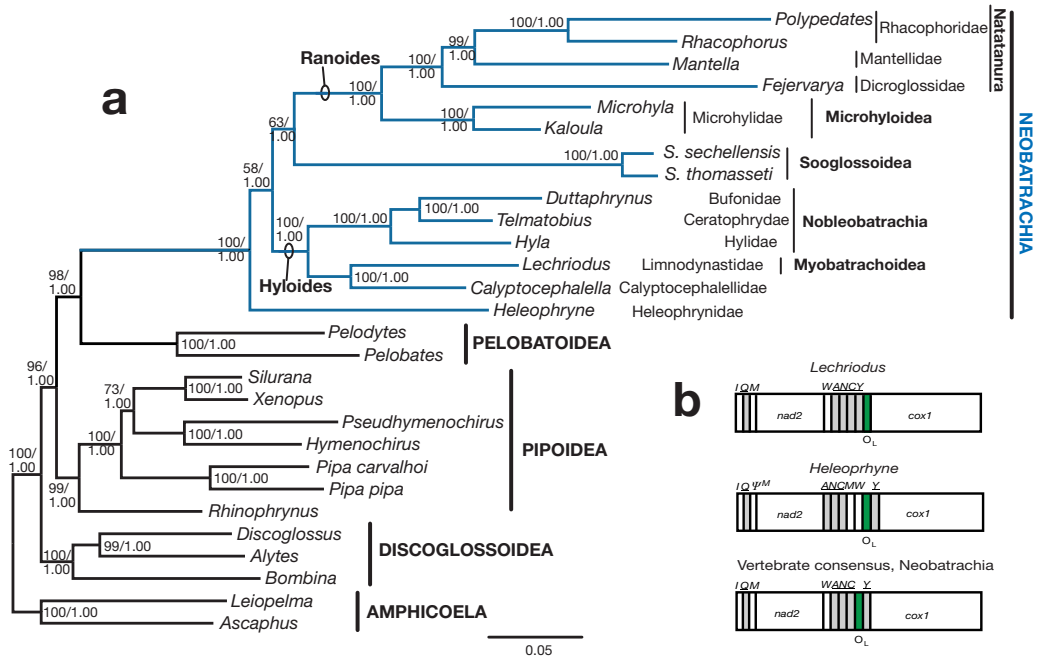


Fig. 4.10. (a) Phylogenetic relationships among frogs (maximum likelihood phylogram) inferred from the combined reduced dataset. Numbers at nodes are support values from maximum likelihood (bootstrap proportions; 1000 replicates, in percent) and Bayesian inference (posterior probabilities). Names of major clades of frogs are shown in capitals, Neobatrachia is highlighted in blue, and familial and supra-familial assignments are indicated for neobatrachians. Scale bar is substitutions · site^{-1}. (b) Derived gene orders in *Heleophryne* and *Lechriodus*, compared to the vertebrate (and neobatrachian) consenus. Genes encoded by the heavy strand are underlined, and tRNA genes are shown by the symbol of their corresponding amino acid.

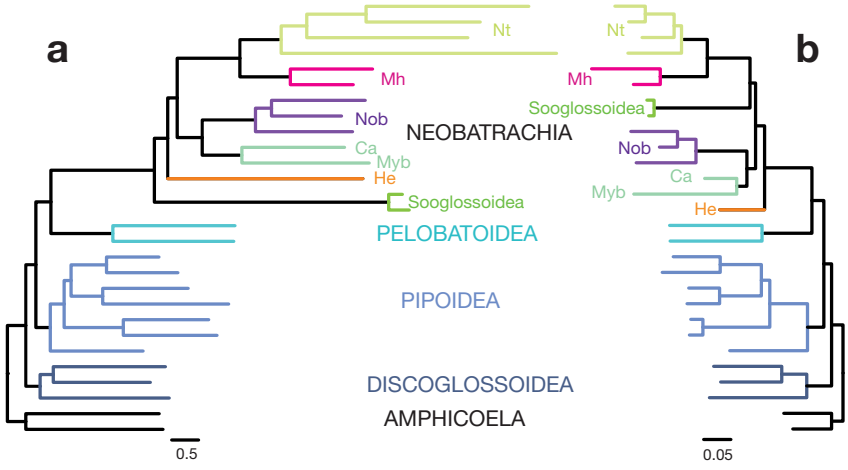


Fig. 4.11. Maximum likelihood phylograms from separate (a) mt and (b) nuclear datasets from the concatenation of the original single-gene alingments. Neobatrachian lineages are abbreviated as Ca, Calyptocephalellidae; He, Heleophrynidae; Myb, Myobatrachoidea; Nob, Nobleobatrachia; Mh, Microhyloidea; and Nt, Natatanura. Both topologies display the same phylogenetic relationships except for the placement of Sooglossoidea. Note that the scale bar (substitutions · site^{-1}) of the mt tree (a) is proportionally 10 times larger than that of the nuclear tree (b), and that neobatrachian branches are distinctively longer in (a) than in (b).

Microhyloidea and Natatanura were recovered as monophyletic, but their relationships with respect to the Afrobatrachia could not be assessed with the available taxon sampling. *Microhyla* and *Kaloula* were grouped into Microhylidae; and within Natatanura, Rhacophoridae was the sister group of Mantellidae to the exclusion of Dicroglossidae.

Phylogenetic analyses of separate mt and nuclear datasets (concatenation of the original single gene alignments with all codon positions of protein-coding genes and non-trimmed rRNA and tRNA alignments; see Materials and Methods 3.4.2 and Table 3.1) rendered two highly congruent topologies with similar levels of support (Fig. 4.11). However, the phylogenetic tree reconstructed based on mt genes (Fig. 4.11a) placed *Sooglossus* as the most basal neobatrachian lineage (branching off before *Heleophryne*), likely due to the attraction of the extremely long branch of the sooglossids by the stem branch of Neobatrachia (which is the longest in the tree) (Felsenstein, 1978a). Additionally, mt genes (Fig. 4.11a) favoured the alternative relationships (*Duttaphrynus* + (*Telmatobius* + *Hyla*)) within Nobleobatrachia, but received lower statistical support (bootstrap proportion = 81%) than in the preferred hypothesis of figure 4.10 or in the nuclear tree (Fig. 4.11b) (bootstrap proportion = 100%; Bayesian posterior probability = 1.00)

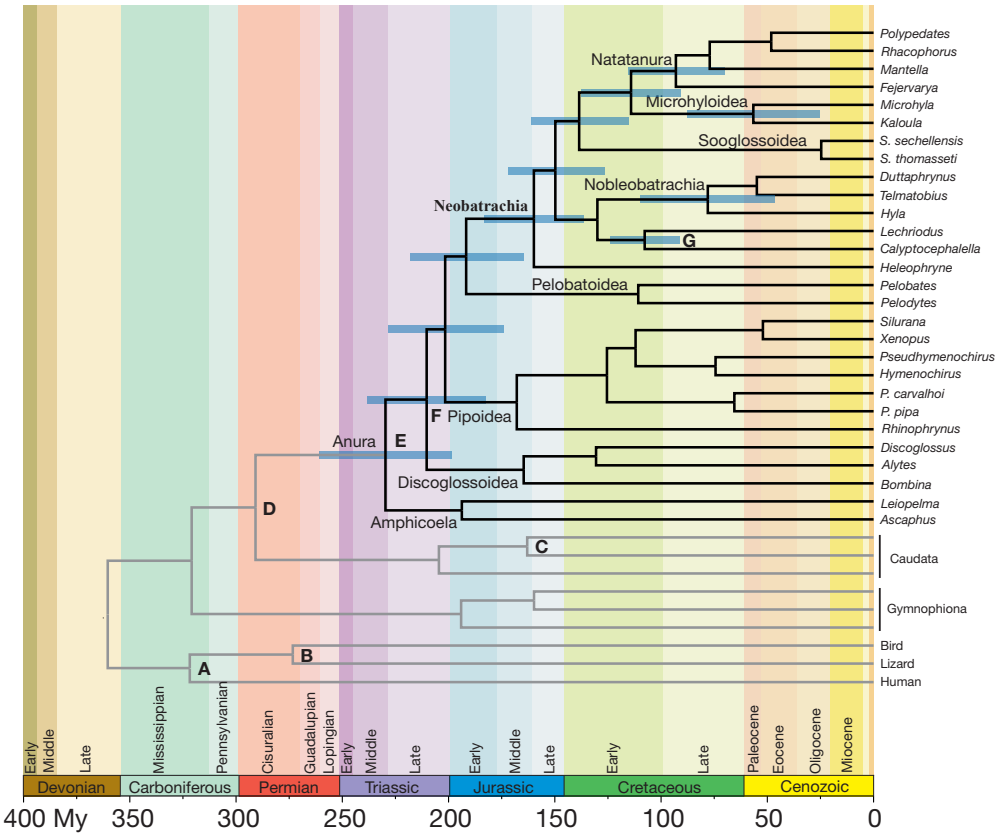


Fig. 4.12. Timetree with age estimates of major divergence events among frogs, based on the combined reduced dataset, and using Bayesian relaxed dating methods (BEAST). Outgroup species are depicted with grey branches, horizontal blue bars represent 95% credibility intervals on relevant nodes (for discussion), and calibration constraints are indicated on the corresponding nodes (A to G) (see main text). Scale is in millions of years.

In addition to assessing congruence, the separate analyses of mt and nuclear genes offers further information on the mode of evolution of these two different genetic systems. Visual inspection of both trees reveals that the branch lengths of the mt tree (Fig. 4.11a), which ultimately correspond to the underlying substitution rates, are more than 10 times longer than those of the nuclear tree (Fig. 4.11b). More importantly, neobatrachians exhibit much longer branches in the mt trees compared to their non-neobatrachian relatives, with *Heleophryne, Sooglossus*, and natatanuran species having the longest branches. Conversely, neobatrachians do not display such conspicuous long branches in the nuclear tree, and the branch lengths do not follow such a clear lineage-specific pattern (Fig. 4.11b).

4.3.3. Estimation of divergence times

The visual inspection of the resulting trees suggests that neobatrachians have distinctively higher substitution rates in their mt genome, which probably became accelerated at the origin of this clade (Fig. 4.10). A molecular clock analysis was performed to provide absolute divergence dates on the particular event (*i.e.,* origin of Neobatrachia), and other major cladogenetic episodes of anuran evolution. The two independent BEAST analyses gave very similar estimates of divergence times for every node (mean difference was 0.64 ± 0.69 million years). To discard the possibility that the estimated dates were determined solely by prior distributions of calibrations, and to confirm the contribution of sequence data, a third BEAST analysis was performed without sequence data (as recommended by the authors in the BEAST documentation), all other parameters being equal. The mean difference of estimated dates was 25.19 ± 25.91 million years.

The origin of crown-group Anura was inferred in the Middle Triassic (~230 mya), and the basal diversification of non-neobatrachian frogs (branching of Amphicoela, Discoglossoidea, Pipoidea, and Pelobatoidea) in the Late Triassic–Early Jurassic (~210-192 mya) (Fig. 4.12). The split between Neobatrachia and Pelobatoidea was dated in the Late Triassic–Early Jurassic, before the initial break-up of Pangaea (mean 192 mya; 95% CI 219-166), and the basal neobatrachian radiation in the Late Jurassic–Early Cretaceous: *Heleophryne* branched off first at 160 mya (95% CI 184-137); and the split between Hyloides and (Ranoides + Sooglossoidea) was dated at 150 mya (95% CI 173-128) (Fig. 4.12). The crown group Nobleobatrachia appeared 79 mya (95% CI 110-47), and Microhyloidea and Natatanura (included in Ranoides) separated 114 mya (95% CI 137-90) (Fig. 4.12).

4.3.4. Lineage-specific substitution rates

In order to test whether the mt substitution rates are significantly higher in neobatrachians (compared to non-neobatrachians), as well as to study lineage-specific rate changes in nuclear genes, a double approach was used, based on (i) relative-rate tests, and (ii) direct comparison of

branch lengths. Furthermore, to deepen into the causes that could have caused such acceleration, (iii) the relative fit of different branch-models was assessed, each assuming different relative rates of replacement and silent substitution (ω) in different branches of the tree, and (iv) it was further examined whether the higher substitution rates were related to a higher number of amino acid synapomorphies in Neobatrachia.

Relative-rate tests clearly showed that neobatrachians had significantly higher mt substitution rates compared to non-neobatrachians (Table 4.5; K2 vs. K1). Mean relative rates of neobatrachians were higher in all mt genes (except for *nad3*), a difference that was significant ($p < 0.05$) in 12 out of the 15 mt genes, and highly significant ($p < 0.001$) when all mt genes were concatenated (Table 4.5). In contrast, some nuclear genes had higher mean substitution rates in neobatrachians (significant only for *rag2* and *slc8a1*), but other genes had lower values (which were not significant) (Table 4.5). When all nuclear genes were concatenated into a single dataset, substitution rates were as a whole significantly higher for neobatrachians than for non-neobatrachians (Table 4.5).

The acceleration of substitution rates could be punctual at the base of Neobatrachia (and thus, not be present in derived lineages), or it could also be a specific feature of the derived and species-rich Ranoides and Nobleobatrachia. Additional relative-rate tests were performed to discriminate among alternative hypotheses regarding the origin of the higher mt rates of neobatrachians: (i) mt substitution rates became accelerated in the origin and were maintained higher along the entire clade; or the higher substitution rates are specific to either (ii) basal or (iii) derived neobatrachian lineages. The comparison of basal (*Heleophryne*, *Calyptochephalella, Lechriodus, Sooglossus*) and derived (Ranoides, Nobleobatrachia) neobatrachians against non-neobatrachian relatives (K3 vs. K1 and K4 vs. K1, respectively) showed that both groups had consistently higher mt substitution rates, which were significant after Bonferroni correction for multiple comparisons ($p < 0.05/ 3 = 0.0167$). However, relative-rate tests of basal versus derived neobatrachians (K3 vs. K4) did not find significant differences in substitution rates (Table 4.5).

On the other hand, nuclear genes did not show a clear pattern when basal and derived neobatrachians were compared separately. After Bonferroni correction for multiple testing, substitution rates in *rag2* gene were significantly higher in both basal and derived neobatrachians compared to non-neobatrachians (K1 vs. K3, and K1 vs. K4, respectively; Table 4.5). For the *slc8a1* gene, the significantly higher rates observed in the previous relative-rate tests (K1 vs. K2; Table 4.5) failed to be significant when basal and derived neobatrachians were separately compared against non-neobatrachians due to the lower significance threshold (Table 4.5). Similar results were also obtained when the same set of relative-rate tests was repeated on amino acid data (Appendix II).

Table. 4. 5. Results from relative rate tests based on nucleotide data of single genes and combined mt and nuclear datasets. Mean weighted substitution rates (K) for (1) non-neobatrachians, (2) all Neobatrachia, (3) basal neobatrachians (*Heleophryne, Calyptocephalella*, *Lechriodus*, and *Sooglossus*), and (4) derived neobatrachians (Ranoides and Nobleobatrachia). Probability values (p) of relative rate tests are shown for each comparison, with corresponding groups in parentheses. Statistically significant results ($p < 0.05$; or $p < 0.5/3 = 0.0167$ after Bonferroni correction) are in bold italics and marked with an asterisk.

Gene	K1	K2	K3	K4	p (1 vs. 2) p<0.05	p (1 vs. 3) p<0.00167	p (1 vs. 4) p<0.00167	p (3 vs. 4) p<0.00167
atp6	0.427863	0.473377	0.481832	0.455958	***0.014730****	***0.005722****	0.105996	0.139169
atp8	0.548420	0.597954	0.603741	0.598698	0.262944	0.238598	0.273487	0.916317
cob	0.357846	0.395422	0.400600	0.376129	0.153630	0.127295	0.461317	0.438624
cox1	0.258230	0.291037	0.294927	0.288984	***1.26·10^{-4}****	***4.86·10^{-5}****	***4.59·10^{-5}****	0.475097
cox2	0.279140	0.349751	0.355794	0.328478	***6.31·10^{-6}****	***3.76·10^{-6}****	***5.41·10^{-4}****	0.068572
cox3	0.289143	0.322240	0.328289	0.304404	***0.007560****	***0.002549****	0.170152	0.046288
nad1	0.371027	0.401329	0.413424	0.383640	***0.023124****	***0.002486****	0.296321	0.020798
nad2	0.507982	0.547513	0.548407	0.525736	***0.025438****	0.026856	0.251499	0.170942
nad3	0.434548	0.430927	0.434119	0.433411	0.869559	0.985402	0.956541	0.974496
nad4	0.431683	0.557581	0.560891	0.590991	***1.00·10^{-7}****	***1.00·10^{-7}****	***1.00·10^{-7}****	0.050825
nad4L	0.497599	0.631415	0.643682	0.588595	***8.27·10^{-4}****	***6.30·10^{-4}****	***0.009870****	0.150316
nad5	0.411432	0.519950	0.517924	0.538763	***1.00·10^{-7}****	***1.00·10^{-7}****	***1.00·10^{-7}****	0.081618
nad6	0.463362	0.505592	0.512071	0.518343	0.088626	0.056897	0.026638	0.765287
rrnS	0.216079	0.289280	0.295438	0.274895	***2.90·10^{-7}****	***1.00·10^{-7}****	***2.32·10^{-6}****	0.078830
rrnL	0.294804	0.323552	0.328272	0.315641	***0.002643****	***8.34·10^{-4}****	0.028736	0.156217
all tRNAs	0.319474	0.362855	0.367254	0.047780	***4.48·10^{-4}****	***1.60·10^{-4}****	0.004985*	0.149114
all mt genes	0.153473	0.194054	0.195809	0.190929	***1.00·10^{-7}****	***1.00·10^{-7}****	***1.00·10^{-7}****	0.141469
bdnf	0.162298	0.183220	0.183143	0.190929	0.063971	0.066639	0.019606	0.425053
cxcr4	0.305898	0.310540	0.310616	0.317277	0.749761	0.760622	0.463425	0.635922
h3a	0.072185	0.071813	0.074309	0.069597	0.967280	0.831715	0.787527	0.662899
pomc	0.383197	0.374114	0.370759	0.381447	0.747452	0.671607	0.956266	0.667038
rag1	0.278115	0.298226	0.297262	0.300048	0.177027	0.210499	0.172426	0.837948
rag2	0.436022	0.493728	0.494946	0.512607	***0.002539****	***0.003285****	***2.38·10^{-4}****	0.314455
rho	0.177431	0.164405	0.167539	0.177528	0.444978	0.596616	0.995982	0.609647
slc8a1	0.236296	0.253733	0.253592	0.255299	***0.029208****	0.041528	0.030535	0.840326
slc8a3	0.223439	0.233366	0.232416	0.238298	0.280968	0.356872	0.115003	0.460431
all nuclear genes	0.067158	0.075283	0.074880	0.072843	***0.036588****	0.055309	0.164760	0.607197

In order to compare the branch-specific rate bias of mt versus nuclear data between neobatrachians and non-neobatrachians, ratios between mt and nuclear branch lengths were calculated and subjected to one-way ANOVA after being log-transformed to meet the assumptions of normality (Shapiro-Wilk's test on residuals $p = 0.062 > 0.05$) and homogeneity of variance (Levene's test $p = 0.947 > 0.05$). The ANOVA analysis showed that log-ratios of branch lengths corresponding to basal neobatrachians, derived neobatrachians, and non-neobatrachians were significantly different ($p = 0.44 < 0.05$). Further orthogonal contrasts found that the ratios were highly significant when basal ($p << 0.001$) and derived ($p = 0.003$) neobatrachians were compared against non-neobatrachians, but they were not significant between basal and derived neobatrachian groups ($p = 0.164$).

The calculation of selective coefficients (ω) for the whole tree of anurans (null model) gave values well below 1 for all mt and nuclear genes (0.005-0.16), indicating the action of purifying selection to maintain gene function (Castellana et al., 2011). To understand whether the observed acceleration of the mt substitution rate in neobatrachians is due to changes in the selective pressure, and to compare the strength of selection acting on the mt and nuclear genomes, we tested for putative changes in the selective regime in four different scenarios for the Neobatrachia (see Materials and Methods 3.7.4). All outcomes are available in the Appendix III, and main results are highlighted in the text below.

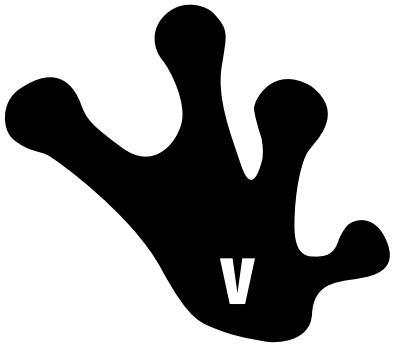
For all mt genes, the independent ω values inferred for the stem branch of Neobatrachia were always higher than those estimated for the whole tree (null model). However, these differences were significant (LRT $p < 0.05$) only for *cob, cox1, cox3,* and *nad1* genes, and for the combination of all mt genes. The independent ω values estimated under alternative models for (i) all neobatrachian branches, (ii) nobleobatrachians, and (iii) ranoideans, were generally higher than those of the null model, but unlike the model of the stem branch Neobatrachia, ω was not higher for every mt gene, and fewer turned out to be significant. These results suggest that purifying selection acting on mt proteins could have been relaxed in neobatrachians.

In order to understand the relative support of the first four models tested, and to further investigate the relevance of the obtained outcome, we compared all 10 models using the AIC (see Materials and Methods 3.7.4). For the combination of all mt genes, the model of relaxed selection in the stem branch of Neobatrachia was clearly better than the remaining models. All other models showed a difference of AIC values (ΔAIC) higher than 10: (i) ΔAIC=48 for the second best model (independent ω for Pipoidea), (ii) ΔAIC=56 for the third (independent ω shared by all neobatrachian branches), (iii) ΔAIC=61 for the fourth (independent ω for Ranoides), etc. A notable exception to the above pattern was the *cox1* gene, because in spite of the overall evidence of relaxed selection in the stem branch of Neobatrachia, the comparison of all 10 models for this gene strongly favoured the relaxation along all branches of Neobatrachia (ΔAIC to the rest of models, including the one assuming an independent ω in the stem branch of Neobatrachia, was >10, and up to 44).

For nuclear genes, most of the estimated independent ω values in all of the nine alternative models were lower than those of the null model, showing evidence of stronger purifying selection, although genes did not display a concordant pattern neither under particular models nor for specific genes. However, there are two exceptions: (i) under the "all neobatrachians" alternative model, relaxation of purifying selection on nuclear genes was recovered in six out of nine genes, even though it was statistically significant only for the genes *rag1, rho*, and *slc8a1*, and the combination of all nuclear genes (LRT $p < 0.05$); (ii) under the model of an independent ω for Amphicoela, for which relaxation of selection was also frequent (in all genes except for *h3a*, although the differences

were only statistically significant for the genes *pomc*, *slc8a1,* and *slc8a3*, and the combination of all nuclear genes). Using the concatenation of all nuclear genes, the comparison of the models assuming a second independent ω for (i) all neobatrachians and (ii) for Amphicoela favoured the latter (ΔAIC=14).

In agreement with the results from the relative-rate tests, which revealed higher substitution rates in neobatrachians, most of the identified amino acid synapomorphies correspond to this clade, although this difference was only significant (binomial test's $p < 0.05$) for the genes *cox1*, *nad5* and *rag2* (see Appendix IV). To further understand how proteins of neobatrachians have accommodated the corresponding mutations, it was investigated whether the synapomorphic amino acids showed any particular pattern of exposition to solvent, or whether they were associated with specific domains of trans-membrane proteins. Results indicated that the distribution of neobatrachian synapomorphic changes were not related to these functional traits, suggesting that mutations were distributed in a more or less uniform manner along mt proteins.



DISCUSSION

5.1. Phylogenetic relationships among frogs

Anura has long been acknowledged as a monophyletic taxon among living amphibians based on both morphological (Rage and Janvier, 1982; Milner, 1988; Benton, 1990; Trueb and Cloutier, 1991; Haas, 2003; Marjanović and Laurin, 2007; Ruta and Coates, 2007; Pyron, 2011) and molecular studies (Zardoya and Meyer 2001; Frost et al., 2006; Pyron and Wiens, 2011). This monophyly is corroborated further by the results of the first study, with strong support from both maximum likelihood and Bayesian inference (Fig. 4.2). In the third study, the origin of crown-group Anura was dated in the Middle Triassic (~ 230 mya; Fig. 4.12), confirming previous time estimates (San Mauro, 2010; Roelants et al., 2011).

5.1.1. Amphicoela and the root of the frog tree of life

Salamanders are the closest living relatives of anurans ("Batrachia" hypothesis; Zardoya and Meyer, 2001; Frost et al., 2006; Carroll, 2007; Ruta and Coates, 2007), and thus, they are generally used to root molecular phylogenies of frogs (San Mauro et al., 2004a; Gissi et al., 2006). Nevertheless, previous studies using mt data pointed out to the problems associated with this decision (García-París et al., 2003b; San Mauro et al., 2004a; Gissi et al., 2006). Salamanders are genetically very distant to living anurans, and this together with the presence of high mt substitution rates in neobatrachian frogs (Hoegg et al., 2004) is likely to produce a long-branch attraction effect, rendering phylogenetic inference more difficult (Felsenstein, 1978a; Swofford et al., 1996).

Anuran taxa that exhibit long branches might be erroneously grouped together irrespective of their true phylogenetic relationships, as well as be attracted to the longest branch in the anuran phylogeny, *i.e.*, that connecting the ingroup (frogs) and the outgroup (salamanders). Moreover, short branches in the tree might also be attracted to each other because of the "leftover" similarity of symplesiomorphic states that "eroded" away in rapid-evolving lineages (Fuellen et al., 2001), thus exacerbating the problems caused by rate variation among lineages in phylogenetic inference. These inconveniences might be alleviated by a detailed study of the extent and distribution of rate variation among frog lineages in both mt and nuclear genomes, as well as by the use of a closer outgroup to root molecular phylogenies of frogs. Given the problematic associated with salamanders, extant members of early branches of the anuran tree might constitute a more appropriate (genetically closer) outgroup taxa for future works interested in inferring robust phylogenetic relationships among frogs.

The genera *Ascaphus* and *Leiopelma* were generally considered the most basal anurans (Duellman and Trueb, 1986), even though their relative phylogenetic positions remained controversial (Lynch, 1973; Duellman and Trueb, 1986; Ford and Cannatella, 1993; Maglia et al., 2001). The mt genome sequence of *Leiopelma archeyi* (together with the one previously available of *Ascaphus*) allowed to unambiguously establish the basal position of *Leiopelma* + *Ascaphus* among Anura ("Amphicoela" hypothesis), receiving strong support from both maximum likelihood and Bayesian inference analyses (Fig. 4.2). This result confirmed previous morphological (Lynch, 1973; Green et al., 1989) and molecular studies (Roelants and Bossuyt, 2005; San Mauro et al., 2005; Frost et al., 2006; Roelants et al., 2007; Blackburn et al., 2010; San Mauro, 2010; Pyron and Wiens, 2011), and rejected the hypothesis of Ford and Cannatella (1993), who placed *Ascaphus* as the sister group of all other remaining frogs including *Leiopelma* (Leiopelmatanura). Phylogenetic relationships at the base of anurans have been rather elusive for traditional morphology-based studies due to the fact that available characters for both *Ascaphus* and *Leiopelma* were either symplesiomorphic or autapomorphic (Green and Cannatella, 1993). However, in the light of our results, the secondary loss of the columella (Stephenson, 1951) and the tail-wagging muscles (Ritland, 1955) may represent morphological synapomorphies supporting the *Leiopelma* + *Ascaphus* clade (Amphicoela).

The hypothesis of Ford and Cannatella (1993), as well as alternative placements of Pipoidea (Maglia et al., 2001; Púgener et al., 2003) or *Leiopelma* alone at the base of the anuran tree were rejected by the AU tests (Table 4.1). Therefore, using *Leiopelma* and *Ascaphus* as outgroup taxa might help to avoid the spurious effects of long-branch attraction, and thus represent an adequate outgroup to root further molecular phylogenies interested in establishing robust phylogenetic hypotheses among main anuran lineages. The timetree analysis estimated the divergence time between *Ascaphus* and *Leiopelma* to have occurred in the Early Jurassic (Fig. 4.11). This is consistent with the suggestion that this split may have been triggered by the early fragmentation of Pangaea (Roelants and Bossuyt, 2005), and easily explains the current disjoint patterns of distribution of these two genera of basal living frogs.

5.1.2. The major lineages of frogs

Five major lineages were recognized within Anura. Non-neobatrachian lineages branched off successively as (i) Amphicoela, (ii) Discoglossoidea (or Costata), (iii) Pipoidea (or Xenoanura), and (iv) Pelobatoidea (or Anomocoela), which was the sister group of (v) Neobatrachia. In the first study, the phylogenetic position of the clades Pipoidea and Discoglossoidea remained ambiguous (they were sister taxa, branching off after Amphicoela; Fig. 4.2). This hypothesis is rather unconventional (also reported in Gissi et al., 2006) because it disagrees with the two most widely accepted hypotheses: (i) (Pipoidea + (Discoglossoidea + (Pelobatoidea + Neobatrachia))) (Haas, 2003; San Mauro et al., 2005; Frost et al., 2006) and (ii) (Discoglossoidea + (Pipoidea + (Pelobatoidea + Neobatrachia))) (Roelants and Bossuyt, 2005; Roelants et al., 2007). In fact, AU tests could not discriminate among the three hypotheses (Table 4.1), as in previous studies (Roelants and Bossuyt,

2005). In the second study, the addition of new mt genome and nuclear data for pipoids, as well as the use of a closer outgroup allowed to resolve this controversial branching order between Discoglossoidea and Pipoidea (Figs. 4.3 and 4.10). The reconstructed phylogeny is congruent with some previous morphology-based (Lynch, 1973; Haas, 1997) and more recent molecular studies (Roelants and Bossuyt, 2005; Roelants et al., 2007; Blackburn et al., 2010; Pyron and Wiens, 2011), but for the first time, alternative branching hypotheses could be ruled out (Table 4.3).

The genera *Pelobates* and *Pelodytes* were found to be sister taxa with high support from both maximum likelihood and Bayesian inference (Figs. 4.10 and 4.11), forming the Pelobatoidea. This lineage was supported as the sister group of Neobatrachia in all maximum likelihood and Bayesian inference analyses with high support (Figs. 4.2, 4.3, 4.10 and 4.11). Both phylogenetic hypotheses agree with most recent studies (Roelants and Bossuyt, 2005; Frost et al., 2006; Gissi et al., 2006; Roelants et al., 2007; Pyron and Wiens, 2011). However, it is in contrast with a sister group relationship between Pipoidea and Pelobatoidea ("Mesobatrachia" hypothesis; Ford and Cannatella, 1993; García-París et al., 2003b), a relationship firmly rejected by topology tests (Tables 4.1 and 4.3).

According to the analyses of divergence times (Fig. 4.12), non-neobatrachian frogs (successive branching of Amphicoela, Discoglossoidea, Pipoidea, and Pelobatoidea) diversified in the Late Triassic–Early Jurassic (~ 210-192 mya; Fig. 4.12). These estimates are in line with other recent molecular dating studies (San Mauro, 2010; Roelants et al., 2011). The obtained divergence times both for the diversification of main anuran lineages, as well as for the splits therein, are especially in agreement with a recent BEAST analysis (Roelants et al., 2011), despite the differences in taxon sampling, choice of molecular markers, and calibration points. However, the obtained time estimates were usually younger than those obtained by earlier studies that used MultiDivTime (Thorne et al., 1998; Thorne and Kishino, 2002) (San Mauro et al., 2005; Zhang et al., 2005a; Roelants et al., 2007), even though the 95% CI mostly overlapped. This discrepancy could be due to differences between the two programs in methodological assumptions of rate change (auto-correlated in MultiDivtime, uncorrelated in BEAST), implementation of evolutionary models and prior calibrations, or techniques for calculating credibility intervals (San Mauro and Agorreta, 2010).

The above results reject the traditional groupings of "Archaeobatrachia" and "Mesobatrachia", long held by taxonomic classifications (Duellman, 1975; Laurent, 1979; Dubois, 1985) and supported by early studies based on mt rRNA genes (Hedges and Maxson, 1993; Hay et al., 1995) (Tables 4.1 and 4.3). The paraphylies of both "Archaeobatrachia" and "Mesobatrachia" have been corroborated by many studies (Frost et al., 2006; Gissi et al., 2006; Pyron and Wiens, 2011). Therefore, the monophyly of "Archaeobatrachia" supported by Hedges and Maxson (1993) and Hay et al. (1995) is likely spurious in the light of all recent molecular evidence, and probably due to the limited data set and long-branch attraction artefacts produced by the presence of among-lineage rate heterogeneity (Roelants and Bossuyt, 2005).

The paraphyly of "Archaeobatrachia" is tightly connected to the hypothesis of vicariance between "archaeobatrachian" and neobatrachian frog families (as well as a vicariance between salamanders and caecilians) produced by the initial break-up of Pangaea in the Mesozoic (Feller and Hedges, 1998). According to these authors, such hypothesis more easily explains the patterns of distributions of neobatrachians in Gondwanan-derived landmasses and non-neobatrachians in Lauransian-derived ones. However, dispersal was invoked to explain two major discrepancies (Feller and Hedges, 1998): (i) the disjoint distributions of *Ascaphus* and *Leiopelma*, and (ii) the distribution of pipoids exclusively in two Gondwanan-derived continents, Africa and South America. The vicariance between "Archaeobatrachia" and Neobatrachia is firmly rejected by the data here presented (Figs. 4.2, 4.3, 4.10, and 4.11, and Tables 4.1 and 4.3) and by most morphological and molecular studies (Haas, 2003; Púgener et al., 2003; San Mauro et al., 2004a; Roelants and Bossuyt, 2005; San Mauro et al., 2005; Roelants et al., 2007; Pyron and Wiens, 2011).

Furthermore, divergence time estimates (Fig. 4.12) clearly point out to an earlier origin of anurans, with initial splits of main lineages of frogs predating the fragmentation of Pangaea, in agreement with recent estimations (San Mauro et al., 2005; Roelants et al., 2007; San Mauro, 2010). In contrast to the hypothesis of Laurasia as the centre of diversification for "archaeobatrachians" and Gondwana for neobatrachians (Feller and Hedges, 1998), a more or less widespread distribution of primitive anurans throughout Pangaea (Bossuyt and Roelants, 2009) is consistent with the reported molecular dating, and it is reinforced by the fossil record of stem group anurans that have been recovered in both Laurasian and Gondwanan-derived land masses (Estes and Reig, 1973; Savage, 1973; Duellman, 1975; Rage and Roček, 1989; Shubin and Jenkins, 1995; Evans and Borsuk-Bialynicka, 1998). In addition, a widespread distribution of primitive anurans throughout Pangaea more easily explains current distribution patterns of *Leiopelma, Ascaphus* and pipoids.

5.1.3. Discoglossoidea

Within Discoglossoidea, most phylogenetic analyses were congruent and recovered a sister group relationship of *Discoglossus* and *Alytes* to the exclusion of *Bombina* (Figs. 4.3 and 4.9), confirming recent morphological (Púgener et al., 2003) and molecular (San Mauro et al., 2004a; Frost et al., 2006; Blackburn et al., 2010; Pyron and Wiens, 2011) phylogenetic studies. These relationships were poorly resolved in the first analysis (Fig. 4.2) probably due to the large number of positions removed from the data matrix and the thinner taxon sampling. Indeed, the latter did not allow testing for the phylogenetic position of *Barbourula*, which has generally been considered sister taxa to *Bombina* (San Mauro et al., 2004a), and supported by molecular data (Blackburn et al., 2010; Pyron and Wiens, 2011).

5.1.4. Pipoidea

The phylogenetic analyses of both mt and nuclear datasets (Figs. 4.3, 4.10, and Table 4.4) support the monophyly of dactylethrines (*Xenopus* and *Silurana*) as in other molecular (Roelants and Bossuyt, 2005; Roelants et al., 2007) and recent morphological studies (*e.g.,* Trueb et al., 2005), and significantly rejected the previously proposed paraphyletic basal arrangement of *Xenopus* and Silurana (Cannatella and Trueb, 1988a, b). The recovered basal position of *Pipa*, and the sister group relationship of dactylethrines and hymenochirines (*Hymenochirus* + *Pseudhymenochirus*) agrees with some previous molecular studies (Roelants and Bossuyt, 2005; Roelants et al., 2007; Pyron and Wiens, 2011). However, it is in stark contrast with a recent phylogenomic study (Bewick et al., 2012), and remarkably, with earlier analyses based on morphology (Cannatella and Trueb, 1988a, b; Púgener et al., 2003; Trueb et al., 2005) that supported a Pipinae clade (including hymenochirines + *Pipa*) with up to six osteological characters (after excluding fossil taxa; Trueb et al., 2005). All alternative hypotheses could be rejected by our molecular dataset, with the exception of a close relationship between *Pipa* and hymenochirines (Table 4.3).

The phylogenomic analysis of Bewick et al. (2012) was based on a comprehensive matrix (missing data was minimal) containing 115 autosomal loci and mt rRNA data (> 35 Kb) for five pipoid species: *Rhinophrynus dorsalis, Xenopus. laevis, Silurana tropicalis, Pipa carvalhoi* and *Hymenochirus curtipes*. Notably, this study lacks information for the pipid genus *Pseudhymenochirus*. A concatenated Bayesian inference approach gave maximal support for a Pipinae clade (Bayesian posterior probability was 1.00), and it was also supported by most analyses of individual data partitions. Interestingly, coalescent-based multilocus species trees agreed with the sister group relationship between *Pipa* and *Hymenochirus*, receiving 0.71-0.86 posterior probability from *BEAST analyses (Heled and Drummond, 2010) and 0.95 from the analysis performed with BEST (Liu and Pearl, 2007). Overall, Bewick et al. (2012) found general congruence among different analyses and settings, but the support evidencing a Pipinae clade is not as strong as could be expected from such a long alignment. In fact, the complexity of this particular phylogenetic question was highlighted by the authors, suggesting that phylogenetic inference might have been confounded by (i) the presence of very large and structured ancestral populations (the distribution of ancestral pipoids spanned much of North and South America, Europe, Middle East, and Africa, Trueb et al., 2005); (ii) the presence of gene flow during speciation (which seems likely given that allopolyploidization between *Xenopus* and *Silurana* is known to occur, but it is not assumed by neither *BEAST nor BEST); or (iii) non-neutral evolution or undetected paralogy in some loci.

If the tree presented in Fig. 4.3 is correct, homoplasy of the six osteological characters supporting a Pipinae clade (Trueb et al., 2005) needs to be assumed. Polarization of these characters is complicated by the fact that all extant pipids are aquatic whereas their unambiguous extant sister group, *Rhinophrynus*, is a terrestrial species with specialized burrowing habits (Duellman and Trueb, 1986). Future morphological studies should assess additional external characters, tadpole morphology, and soft anatomy in the search for possible synapomorphies of the four African genera, such as the keratinization of the first three toes, which is more strongly expressed in the African taxa (Dunn, 1948).

From a biogeographic point of view our hypothesis suggests that the basal split among extant pipids might have separated an African lineage (*Hymenochirus, Pseudhymenochirus, Silurana,* and *Xenopus*) from a South American lineage (*Pipa*) and is consistent with the American distribution of Rhinophrynidae as sister group of the Pipidae (Noble, 1931; Dunn, 1948). This indicates the need of re-evaluating also the phylogeny of fossil taxa, given the apparent biogeographic anomaly that the South American *Pipa* based on morphological data is nested within a clade of purely African fossil taxa, and the African *Silurana* + *Xenopus* are placed within a clade of exclusively South American fossil taxa (Trueb et al., 2005).

5.1.5. Pelobatoidea

The genera *Pelobates* and *Pelodytes* are sister taxa, corroborating the Pelobatoidea, as in most morphological (Duellman and Trueb, 1986; Ford and Cannatella, 1993; Púgener et al., 2003) and molecular studies (García-París et al., 2003b; Frost et al., 2006; Pyron and Wiens, 2011), but contrasting with the phylogenetic analysis based on larval morphology of Haas (2003), who found *Pelodytes* more closely related to *Heleophryne* than to the rest of Pelobatoidea. The taxon sampling was insufficient to provide a general overview of the phylogenetic relationships among all four recognized families. Recent studies support either Scaphiopodidae as sister group of (Pelodytidae + (Pelobatidae + Megophryidae) (García-París et al., 2003b; Pyron and Wiens, 2011) or a sister group relationship between (Pelodytidae + Scaphiopodidae) and (Pelobatidae + Megophryidae) (Frost et al., 2006).

5.1.6. Neobatrachia

Neobatrachia has traditionally been acknowledged to be monophyletic (Reig, 1958; Duellman, 1975), a fact that has been corroborated by our analyses (Fig. 4.2, 4.3, 4.10, and 4.11) and most morphological (Duellman and Trueb, 1986; Ford and Cannatella, 1993) and molecular studies (Hoegg et al., 2004; Roelants and Bossuyt, 2005; San Mauro et al., 2005; Frost et al., 2006; Roelants et al., 2007; Pyron and Wiens, 2011). The split between Neobatrachia and Pelobatoidea was dated in the Late Triassic–Early Jurassic, before the initial break-up of Pangaea (mean 192

mya; 95% CI 219-166) (Fig. 4.12). The obtained phylogenetic tree recovered *Heleophryne* as the sister group to all other neobatrachians (Fig. 4.11), as reported by recent molecular studies (Hoegg et al., 2004; Frost et al., 2006; Roelants et al., 2007; Pyron and Wiens, 2011), but it received high support only from Bayesian inference (Fig. 4.10). Other studies have suggested a sister group relationship of *Heleophryne* with Myobatrachoidea (Biju and Bossuyt, 2003) or with Myobatrachoidea + Sooglossoidea (Duellman and Trueb, 1986). Moreover, Haas (2003) placed *Heleophryne* within Pelobatoidea, rendering Neobatrachia paraphyletic. Papuan-New Guinean *Lechriodus* and South American *Calyptocephalella* were recovered as sister genera (Fig. 4.10), and both as sister to Nobleobatrachia. These results point to a close relationship of *Calyptocephalella* to at least some components of Limnodynastidae, as found by recent molecular analyses (San Mauro et al., 2005; Frost et al., 2006; Roelants et al., 2007; Pyron and Wiens, 2011), and contrary to the traditional grouping of *Telmatobius* and *Calyptocephalella* in the family "Leptodactylidae" (Lynch, 1971; Lynch, 1973).

According to the divergence time estimates, the basal neobatrachian radiation occurred in the Late Jurassic–Early Cretaceous: *Heleophryne* branched off first at 160 mya (95% CI 184-137); and the split between (Nobleobatrachia + (Myobatrachoidea + *Calyptocephalella*)) and (Ranoides + Sooglossoidea) was dated at 150 mya (95% CI 173-128), in agreement with divergence time estimates of other recent studies (Vences et al., 2003; San Mauro et al., 2005; Zhang et al., 2005a; Igawa et al., 2008; Roelants et al., 2011). However, these estimates are much older than the oldest known neobatrachian fossils (calyptocephalellid fossils from Late Cretaceous; Sanchíz, 1998; Báez et al., 2000; Roček, 2000; Roček and Rage, 2000). Thus, the current fossil record for Neobatrachia is a poor indicator of the particular split between Pelobatoidea and Neobatrachia. The separation between Nobleobatrachia and (Myobatrachoidea + *Calyptocephalella*) occurred 79 mya (95% CI 110-47), and Microhyloidea and Natatanura (included in Ranoides) separated 114 mya (95% CI 137-90), in agreement with other recent studies (Zhang et al., 2005a; Igawa et al., 2008; Roelants et al., 2011).

Within Nobleobatrachia, *Duttaphrynus* (Bufonidae) and *Telmatobius* were sister taxa to the exclusion of *Hyla*, in agreement with some molecular studies (Frost et al., 2006; Pyron and Wiens, 2011), but contradicting the morphological analyses of Lynch (1971; 1973) and other molecular studies. San Mauro et al. (2005) found *Hyla* and *Bufo* (Bufonidae) to be more closely related, whereas Roelants et al. (2007) found *Telmatobius* and *Hyla* to be sister taxa to the exclusion of *Duttaphrynus*. The two species of *Sooglossus* included were grouped together (hence corroborating this genus; van der Meijden et al., 2007a) at the base of Ranoides (Fig. 4.10). However, support for this branch was only moderate (maximal for Bayesian inference, but low for maximum likelihood), as in other recent molecular studies (Hoegg et al., 2004; San Mauro et al., 2005; Roelants et al., 2007). This hypothesis is in line with Savage's (1973) suggestion of a close relationship between sooglossids and ranoids, but contrasts with others who suggested the basal position of Sooglossidae among

"Hyloidea" *sensu lato* (Laurent, 1979; Ruvinsky and Maxson, 1996), or grouped this family with Microhyloidea (Blommers-Schlösser, 1993), with Myobatrachinae inside hyloids (Lynch, 1973; Duellman and Trueb, 1986; Ford and Cannatella, 1993), or with Nasikabatrachidae at the base of neobatrachians (Biju and Bossuyt, 2003; Pyron and Wiens, 2011). Ranoidean intra-relationships received high support (Figs. 4.3 and 4.10): Microhyloidea and Natatanura were recovered as monophyletic, but their relationships to the Afrobatrachia could not be assessed with the available taxon sampling. The genera *Microhyla* and *Kaloula* were grouped together forming the Microhylidae, in agreement with many other molecular studies (*e.g.*, Frost et al., 2006; van der Meijden et al., 2007b). Within Natatanura, Rhacophoridae was grouped with Mantellidae to the exclusion of Dicroglossidae, as in other recent molecular studies (Frost et al., 2006; Roelants et al., 2007; Bossuyt and Roelants, 2009).

Overall, the obtained tree shows a fully resolved and robust phylogeny of frogs (Figs. 4.3. and 4.10), which receives support and agreement from both molecular and morphological evidence. Some uncertainty still remains regarding the phylogenetic position of Sooglossidae (+ Nasikabatrachidae; Biju and Bossuyt, 2003) and Heleophrynidae. Despite the agreement of the obtained tree with other recent studies, support values from maximum likelihood are still low to provide a definitive answer, although Bayesian posterior probabilities are maximal.

5.2. Evolution of sound production in Pipidae

Behavioural observations suggest that sound production in *Pseudhymenochirus* occurs by air movement from the lungs to the throat (Fig. 4.4 and movie, available at http://www.biomedcentral.com/1471-2148/11/114/additional). This is unique among pipids, which generally produce motionless clicking sounds by implosions related to the derived box-like structure of the larynx (Yager, 1992), and in fact more similar to the typical mechanism found in non-pipid frogs. The sound production mechanism of pipids has been thoroughly studied in *Xenopus borealis* (Yager, 1992) and given the resemblance of calls and motionless calling behaviour in other members of the family (Kunz, 2003), it is assumed to be the general system in pipids. In *X. borealis,* the characteristic clicking sound was proved to be produced by the simultaneous action of bipennate muscles that separate the discs of ossified arytenoid rods (Yager, 1992). The sound is produced by the implosion of air when the two arytenoid discs separate, given that no clicking sound was emitted when this space was filled with liquid (Yager, 1992). Similarly, the implosion mechanism is not air-driven, because call spectra remained unchanged after frogs were forced to breathe helium (Yager, 1992).

With regards to morphological analyses, the results are fully congruent with previous detailed anatomical descriptions in which the larynges of pipids are enlarged boxes formed by more or less ossified cartilages (Ridewood, 1897, 1900; Rabb, 1960; Yager, 1996). Despite the apparent diversity in larynx morphology both among pipids and among frogs, the embryological origin of involved cartilages have been traced back to the larval hyobranchial apparatus (Duellman and Trueb, 1986), leaving little doubt of their homology within amphibians (Parker, 1871; Ridewood, 1897).

The larynx of *Hymenochirus* is an enlarged box with ossified cartilages, more similar to that of *Xenopus* and *Silurana* than to other non-pipid anurans (*e.g.,* Bombina in Fig. 4.6). In *Pipa*, although the larynx structure slightly differs from that of *Xenopus* (Yager, 1992, 1996) a similar sound production has also been suggested (Rabb, 1960). The larynx of *Pseudhymenochirus* is particularly similar to that of *Hymenochirus* and shows the typical ossified cartilages of other pipids (Ridewood, 1897, 1900). Therefore, it can undoubtedly be asserted that the larynx in *Pseudhymenochirus* evolved from a typical pipid condition, but the overall structure seems to be more flexible, and this could somehow permit a movement of air that is used to vocalize, as suggested by the performed behavioural observations. However, whether vocal cords are present in *Pseudhymenochirus* (which are absent in pipids) or whether other structures are responsible for sound production requires further specific examination. Other hypotheses may also be plausible, and further detailed functional studies (as those performed by Yager [1992]) are needed in order to determine the exact mechanism through which sound is produced, as well as the precise function of involved structures.

Molecular data leave no doubt of the nested phylogenetic position of *Pseudhymenochirus* within Pipidae (Fig. 4.3 and Table 4.3). This is so even under the alternative hypothesis not rejected by AU tests (*Pipa* as the sister group of hymenochirines; Table 4.3), which is supported by morphological data and a recent phylogenomic study. Moreover, using this latter hypothesis for ancestral character state reconstruction also supoorted the sound production mechanism, as being independent of air in the origin of the family Pipidae. Morphological data clearly demonstrate the pipid nature of the *Pseudhymenochirus* larynx (Fig. 4.8). While the source used for sound production unexpectedly appear to reverse to the ancestral non-pipid condition (movement of the air column), associated anatomy evolved from a typical pipid-like larynx that likely imposed constraints to natural selection. Altogether, it is suggested that the air-driven sound production in *Pseudhymenochirus* most probably represents a novel evolutionary combination and it is a remarkable example of complex anatomical modifications related to a functional shift of enormous influence in frog behaviour and reproduction.

The selective forces for these changes are unknown, but the movements of the body flanks during the call in *Pseudhymenochirus* obviously produce water waves that might provide information about the size of the calling male to females, detected by their lateral line system. Water surface waves can play an important role in the advertisement behaviour of several basal anurans (*e.g.,* discoglossoids; Walkowiak and Münz, 1985; Glaw and Vences, 1991). Compared to *Hymenochirus*, sexually active *Pseudhymenochirus* males have morphologically less distinct postaxillary glands (Fig. 4.4), which are used in chemical communication during the breading season in *Hymenochirus* (Pearl et al., 2000). Therefore, flank movements in *Pseudhymenochirus* could serve as additional visual and mechanical signals, which might reinforce the acoustic signals to attract females and impress conspecific males.

This study exemplifies that understanding the evolutionary process underlying an innovation, here the air-driven call in *Pseudhymenochirus*, can only be achieved with an integrative comparative approach. In this particular case, behavioural observations prompted for detailed anatomical analyses, and comparative data were placed within a robust phylogenetic framework based on molecular data. Further insights on the nature of this evolutionary innovation could be gained through ontogenetic studies that disentangle how morphological constraints imposed by the rather stiff larynx box of pipids are overcome to allow the reversal to the ancestral air-driven vocalization in *Pseudhymenochirus*. This strengthens the idea that evolutionary solutions to functional problems often emerge based on previous structures, and for this reason, innovations largely depend on possibilities and constraints predefined by the particular history of each lineage. The result of this study provides yet another example of how natural selection generates complex morphologies and functions by tinkering with previously available structures (Jacob, 1977), and further reinforces the important roles of historical contingency and constraints in canalizing potential solutions to a given evolutionary problem (Gould, 1980).

5.3. Evolution of mitochondrial genome rearrangements in anurans

The gene content and structure of the mt genome is highly conserved across vertebrates (Boore, 1999; Gissi et al., 2008). Most amphibians display the consensus gene order of vertebrates (Roe et al., 1985; Zardoya and Meyer, 2000, 2001; Zardoya et al., 2003; San Mauro et al., 2004a; Zhang et al., 2005a; Gissi et al., 2006; Pabijan et al., 2008; Zhang and Wake, 2009a, b), even though several exceptions have been reported in caecilians (San Mauro et al., 2006), salamanders (Mueller et al., 2004; Mueller and Boore, 2005) and frogs. In frogs, gene rearrangements had only been reported among neobatrachians (Table 1.2 and Fig. 5.1; Sano et al., 2005; Kurabayashi et al., 2006; Igawa et al., 2008; Kurabayashi et al., 2008), with non-neobatrachian frogs generally

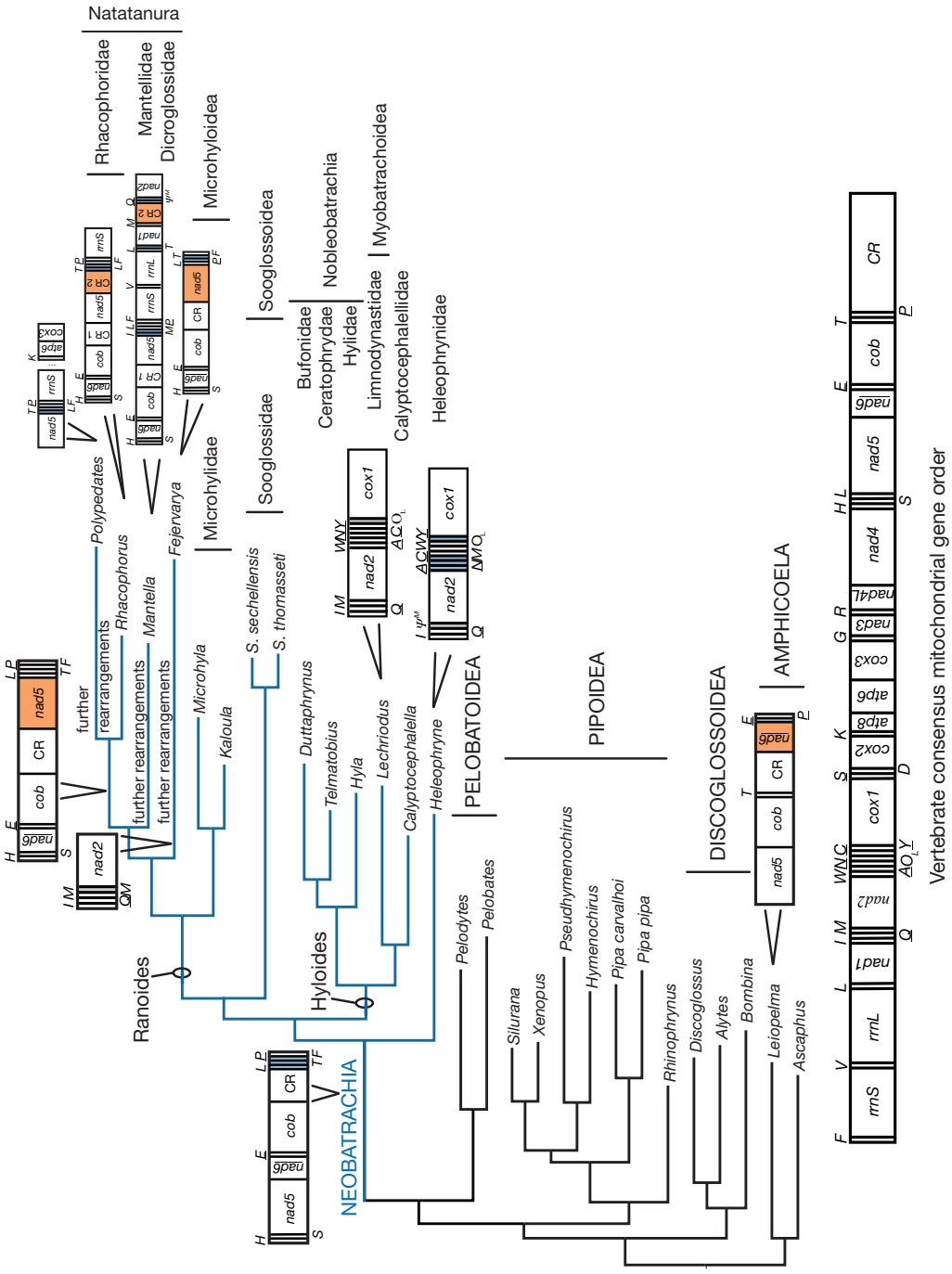


Fig. 5.1. Evolution of mitochondrial gene order in Anura. Changes in the mitochondrial gene order are mapped onto the preferred phylogenetic hypothesis of Fig. 4.10. and the regions with translocated genes (compared to the vertebrate ancestral gene order; lower part) are shown on the corresponding branches. Note that *Fejervarya limnocharis, Mantella madagascariensis,* and *Rhacophorus schlegelii* and *Polypedates megacephalus* do not represent the ancestral gene orders of Dicroglossidae, Mantellidae, and Rhacophoridae, respectively; and rather display more derived gene orders. Further mitochondrial genome rearrangements are known to occur in these families. Genes encoded by the heavy strand are underlined, and tRNA genes are shown by the symbol of their corresponding amino acid. Pseudogenes of the *trnM* genes are shown with ψ^{M}, and the discontinuity between two mitogenomic regions in *Fejervarya* are shown with three points.

conforming to the vertebrate consensus (San Mauro et al., 2004a; Gissi et al., 2006). With the exception of *Leiopelma archeyi* (Fig. 4.1), all other non-neobatrachian frogs whose mt genome has been sequenced in this thesis (including *Rhinophrynus dorsalis, Pipa carvalhoi, Hymenochirus boettgeri, Pseudhymenochirus merlini, Xenopus laevis* and *Pelodytes punctatus*) conform to this consensus order of vertebrates. Hence, *Leiopelma archeyi* is the first reported non-neobatrachian species departing from this consensus; and furthermore, this mitogenomic order is new for frogs, being clearly derived from the ancestral gene order of vertebrates.

Neobatrachian frogs display a variety of mt gene orders, with the common distinctive feature of a *LTPF* tRNA gene cluster between the control region and the *rrnS* gene, produced by the translocation of the *trnL–(CUN), trnT* and *trnP* genes from upstream of the control region (Sumida et al., 2001). However, the lack of mitogenomic information for basal families within Neobatrachia hindered the exact origin of this particular gene order. The newly reported neobatrachian mt genomes generally conform to the typical gene order of neobatrachians described above (*Calyptocephalella gayi, Telmatobius bolivianus, Sooglossus thomasetti* and *S. sechellensis*), with the exception of *Heleophryne regis* and *Lechriodus melanopyga*. *Heleophryne regis* departs from the consensus neobatrachian order in a rearrangement involving the region between the *IQM* and *WANCY* tRNA clusters, which are similarly organized in both neobatrachians and non-neobatrachians. Therefore, the neobatrachian-specific mt gene arrangement (Sumida et al., 2001) can be confidently regarded as the ancestral order for this clade. Therefore, the new gene order found in *H. regis* is derived from the ancestral gene order of neobatrachians. It might be possible that the long intergenic spacer between the *trnP* and *trnF* genes found in *H. regis* could be a remnant of the ancestral tandem duplication and random loss event by which the *LTPF* cluster could have originated in the origin of Neobatrachia (Sumida et al., 2001). The translocation of the origin of replication of the light strand found in *Lechriodus melanopyga* is clearly derived from the ancestral neobatrachian gene order given its unambiguously nested position within Neobatrachia (Fig. 5.1).

Among other neobatrachian species for which mt genome data is available, most of them follow the ancestral gene order of Neobatrachia (Sumida et al., 2001), including *Calyptocephalella*, Nobleobatrachia, *Sooglossus* and Microhyloidea (Fig. 5.1). Most of the reported arrangements that depart from the neobatrachian-specific gene order occur within Natatanura. In the light of the available mitogenomic information, two duplications occurred in the ancestor of all Dicroglossidae: (i) a duplication of *trnM* gene in which both *trnM* gene copies remain adjacent, between the *trnQ* and *nad2* genes, and (ii) a duplication of the region from the 5' end of *nad5* to 3' end of the *trnP* gene. One of the duplicated *trnM* gene copies is secondarily lost in *Limnonectes fujianensis* (Alam et al., 2010) and several gene arrangements can be found in the region between the *nad5-trnP* genes among dicroglossids, probably resulting from the loss of different paralogs in different lineages (Alam et al., 2010). The ancestral gene order of Ranidae seems to conform to the

neobatrachian type, even though further mt genome rearrangements have been reported within this clade (Kurabayashi et al., 2010).

In the common ancestor of Rhacophoridae and Mantellidae, a translocation of the *nad5* gene seems to have occurred, from its ancestral location, between the *trnS–(AGY)* and *nad6* genes, to the new position downstream, between the *cob* and *trnL–(CUN)* genes (Kurabayashi et al., 2006; Kurabayashi et al., 2008). Further gene rearrangements have also been reported within Rhacophoridae and Mantellidae (Sano et al., 2005; Kurabayashi et al., 2008).

The maintenance of pseudogenes in mt genomes is remarkable due to the strong selective pressure for a compact organization (Rand, 2001). However, the presence of *trnM* pseudogenes has repeatedly been observed in species where the *trnM* gene is translocated into a new position, including *Heleophryne regis*, several members of the family Mantellidae (Kurabayashi et al., 2006; Kurabayashi et al., 2008) and a number of phylogenetically distant species of fishes (Mabuchi et al., 2004). In *Heleophryne regis* and in most of the above-mentioned cases, anticodon sequences are mostly degenerated (indicating loss of function), yet pseudogenes retain clover-leaf secondary structures (Mabuchi et al., 2004). Hence, it has been proposed that *trnM* pseudogenes could have retained an active role as a punctuation mark for the precise processing of the 5' end of *nad2* messenger RNA during replication (Mabuchi et al., 2004; Kurabayashi et al., 2006). The hypothesis that tRNA genes could act as punctuation marks (Ojala et al., 1981) is further supported by the identification of precise cleavage at 5' and 3' ends of tRNA genes in the human mt genome (Rossmanith et al., 1995) and tRNA mutations causing maternally transmitted diseases produced by processing defects (Levinger et al., 2003).

All of the observed gene rearrangements can be explained by the tandem duplication—random loss model of gene order change (Moritz and Brown, 1986, 1987; Moritz et al., 1987; Boore, 2000), which is confidently regarded as the dominant mechanism of gene order rearrangement in vertebrate mt genomes (San Mauro et al., 2006). In *Leiopelma archeyi*, an ancestral tandem duplication of the mitogenomic region involving the genes *nad6, trnE, cob, trnT, trnP* and the control region (CR 5' region) could have occurred, followed by the arbitrary loss of redundant gene duplicates. Likewise, the transition from the vertebrate consensus order (common in most non-neobatrachians) to the ancestral rearrangement of Neobatrachia, could have occurred after a single duplication of the region (at least) involving the genes *trnL–(CUN), nad5, nad6, trnE, cob, trnT,* and *trnP* and the control region.

Further support for the tandem duplication—random loss model is provided by (i) the presence of a *trnM* pseudogene in *Heleophryne regis*, which could be a remnant of the process of duplication and random loss, and (ii) the presence of tandem repeats in the non-coding region between the *trnP*

and *trnF* genes in *H. regis.* All other reported rearrangements in the mt genome of anurans are also compatible with the tandem duplication—random loss model (Sumida et al., 2001). However, it has been suggested that a high number of tandem duplication events would be necessary to explain the mt gene rearrangements in mantellids (Kurabayashi et al., 2008) or the transposition of single tRNA genes in ranids (Kurabayashi et al., 2010), suggesting that alternative mechanisms might have occurred in these cases. Such alternative mechanisms include homologous recombination (Amer and Kumazawa, 2007) or the tRNA-priming model (Jacobs et al., 1988), but given that no changes in the sense strand of genes are observed in mantellids (unlike in the lizard *Calotes versicolor*; Amer and Kumazawa, 2007), the tandem duplication—random loss model cannot be completely ruled out.

All reported mt genome rearrangements (including that of *Leiopelma*, the origin of the *LTPF* tRNA gene cluster in Neobatrachia and the derived gene orders of both *Heleophryne* and *Lechriodus*) are associated with origins of replication, which are considered hot spots of gene order change in the vertebrate mt genome (Mindell et al., 1998; San Mauro et al., 2006). Previous studies have indicated that duplications are more likely to take place in close proximity to (or involving) replication origins due to mechanistic constraints (Moritz and Brown, 1986; Mindell et al., 1998; Dowton and Austin, 1999; Boore, 2000). Alternatively, it has been proposed that tRNA genes could act as promoters of gene duplications due to either secondary structure or sequence similarities between different tRNA genes (Moritz and Brown, 1987).

The incorporation of the nascent chain during the replication of the heavy strand has also been proposed as an alternative mechanism that could promote gene rearrangements near the control region (Zardoya et al., 1995b). During the replication of the heavy strand in the mt genome, most newly initiated chains are arrested by the termination-associated sequences (TAS; Doda et al., 1981; MacKay et al., 1986) and their replication, thus, finishes downstream, shortly after the origin of replication of the heavy strand (1350–1510 bp in *Xenopus laevis*; Roe et al., 1985). The newly synthesized chain remains associated with the template, creating a triple-stranded structure known as the D-loop (Clayton, 1982; MacKay et al., 1986; Bowmaker et al., 2003), which may be responsible for gene rearrangements by non-homologous recombination of the nascent chain (Zardoya et al., 1995b). Furthermore, the phenomenon of replication fork arrest is a well-recognized prelude to gene rearrangement in the nuclear genome (Hyrien, 2000; Rothstein et al., 2000; Bidnenko et al., 2002).

Rare genomic changes such as gene rearrangements represent very attractive tools for phylogenetic inference due to the theoretically large space of possible arrangements and thus, low levels of homoplasy (Boore and Brown, 1998; Rokas and Holland, 2000). However, gene duplications, and as a consequence, gene rearrangements, may be more frequent in certain regions (hot spots). Thus, the number of different gene orders that are observed is reduced among

all theoretically possible rearrangements, and the likelihood of convergence is increased (San Mauro et al., 2006). The mt gene order found in *L. archeyi* (Fig. 4.1) is convergent with two species of salamanders and 14 species of conger eels (Table 4.2).

Besides *L. archeyi*, more instances of gene order rearrangements occur in the CR 5' region of vertebrates (Table 4.2). Assuming the tandem duplication—random loss model and a single duplication event, 27 different gene orders are, in principle, possible (see section 4.1.3 of Results). Among currently available mt genomes of vertebrates, seven different arrangements were found. Some of these gene orders were convergent, this seven orders being the product of at least 12 independent origins across the vertebrate phylogeny. The convergences in mt gene order of *L. archeyi* with other eels and salamanders, along with the comparative data for other available vertebrates provide compelling evidence that the 5' end of the control region is a hot spot for gene order rearrangement in vertebrate mitochondria.

5.4. An overall substitution rate shift in Neobatrachia?

The new data added for key basal lineages of Neobatrachia was essential to understand the origin of the higher mt substitution rates in this clade. Both relative-rate tests and topological measures support a statistically significant acceleration of the mt substitution rate shared by both basal and derived lineages within Neobatrachia (Table 4.5), corroborating previous studies that suggested an unequal distribution of mt substitution rates among frogs (Feller and Hedges, 1998; Hoegg et al., 2004; San Mauro et al., 2004a; Gissi et al., 2006; Igawa et al., 2008). Phylogenetic analyses suggest that the origin of this rate acceleration began at the stem branch leading to Neobatrachia, in the Early-Middle Jurassic period (Fig. 4.12).

Nuclear genes did not exhibit a congruent trend of neobatrachian-specific higher substitution rates, confirming an earlier study (Hoegg et al., 2004). Noticeably, the substitution rate for the concatenation of all nuclear loci was significantly higher in neobatrachians than in non-neobatrachians, but this difference was not significant when basal and derived neobatrachians and non-neobatrachians were compared pairwise, and the p threshold corrected for multiple multiple comparisons (Table 4.5). It might be possible that the concatenation of all nuclear genes revealed a hidden common pattern among nuclear genes, which could not be manifested when genes were analyzed individually. However, unlike for mt genes, not all nuclear genes possess neobatrachian-specific higher substitution rates, suggesting that this result could be spurious. Therefore, the significantly higher rates from concatenated nuclear loci might be explained by the desproportionate contribution of particular loci, specially *rag2*, which is clearly accelearated in the Neobatrachia (Table 4.5).

Because substitution rates are determined, to a great extent, by the balance between selection and genetic drift (Bromham, 2009b), changes in synonymous and non-synonymous rates were studied in frogs, and found that purifying selection acting on mt proteins might have been relaxed in Neobatrachia. Among all tested models, the assumption of relaxation at the stem branch leading to Neobatrachia was clearly better than the rest, although the very similar results obtained under the model of a general relaxation along the entire Neobatrachia indicates that this alternative hypothesis cannot be confidently rejected. These changes in the selection coefficient could explain, at least in part, why mt substitution rates became accelerated in the origin of Neobatrachia, as found by relative-rate tests and topological measures. Nonetheless, the strong evidence for the relaxation of selection in the *cob* bene along the stem of Neobatrachia (LRT $p = 0.003 << 0.05$) contrasts with the non-significance of the corresponding relative-rate tests, although substitution rates were higher in neobatrachians (although K2, K3, K4 > K1; Table 4.5). The interpretation of the results from nuclear genes is more complex, and remarkably, the suggested relaxed selection along all neobatrachian branches does not explain the higher substitution rates found by relative-rate tests for the *rag2* gene, although it does for the *slc8a1* gene.

Many of the comparative studies of molecular evolutionary rates carried out to date have been primarily based on the assumption of a systematic component for rate variation, trying to find a correlation with a given trait, which could eventually explain the observed variation in rates (Bromham, 2009a). Some studies found a correlation between higher mt substitution rates and more events of gene rearrangements in the mt genome of some metazoans (Shao et al., 2003; Xu et al., 2006). However, because most gene rearrangements among frogs occur mostly within Natatanura (Fig. 5.1; Liu et al., 2005; Sano et al., 2005; Ren et al., 2009; Kurabayashi et al., 2010), and because mt substitution rate became accelerated well before (origin of Neobatrachia), frogs do not appear to conform to this pattern. Furthermore, Kurabayashi et al. (2008) found evidence for the absence of correlation between mt rates and number of gene rearrangements in one intensively studied lineage of neobatrachians (mantellid frogs from Madagascar).

Many other studies have found substitution rates to be correlated with species diversification (Barraclough and Savolainen, 2001; Webster et al., 2003; Eo and DeWoody, 2010; Lanfear et al., 2010), and three main hypotheses have been proposed to explain this correlation (Lanfear et al., 2010). (i) Increased diversification in certain lineages could promote higher substitution rates because speciation is often associated with processes that can potentially increase substitution rates, such as adaptation to new environments or transient reductions in population sizes (Pagel et al., 2006; Venditti and Pagel, 2009). (ii) Higher substitution rates could produce higher net diversification, both by increasing speciation rate and/ or by reducing extinction rate (Lanfear et al., 2010). Species with shorter generation times or higher metabolic rates can be expected to have higher genetic diversity within populations due to higher mutation rates, thus making fixations

more likely (Bromham et al., 1996; Bromham and Leys, 2005). The impact of mutation rates on substitution rates remains controversial (Lanfear et al., 2007), but lineage-specific differences in mutation rates might be especially important in populations under non-equilibrium conditions, after *e.g..* bottleneck events (Nabholz et al., 2008). (iii) A third hypothesis, however, rejects a causal relationship between substitution and diversification rates, and holds that this correlation is indeed due to other factors that influence both; for example, environmental energy in the case of the diversification of flowering plants (Davies et al., 2004).

In frogs, it has been suggested that the observed higher mt substitution rates of neobatrachians could have been the product of faster recent speciation events in this clade (and thus more bottleneck events) (Hoegg et al., 2004). In addition, Dubois (2004) hypothesized that direct-developing species (mostly within Neobatrachia) would tend to have higher substitution rates, and this in turn would promote speciation. Frogs with direct development lay eggs in clutches, and thus they are subjected to "familial" mortality, in contrast to "individual" mortality in species with tadpole stage, increasing the probability of fixation of alleles, and ultimately promoting higher diversification rates (Dubois, 2004). Neobatrachians do not fit the hypothesis of the different reproductive modes (Dubois, 2004) because, although basal neobatrachian lineages are mostly indirect-developers and most direct-developers belong to derived lineages within Neobatrachia (Dubois, 2004), both groups display higher mt substitution rates (Table. 4.5).

Alternatively, it has also been suggested that higher substitution rates of neobatrachians could be responsible for higher diversification rates, due to shorter generation times and/ or higher metabolic rates (Hoegg et al., 2004). Both basal and derived lineages of neobatrachians share higher mt substitution rates compared to non-neobatrachian relatives (Table 4.5), but species diversity is highly unequally distributed among them, with most of the diversity corresponding to derived lineages, and basal neobatrachians represented by few species with relict distribution (Table 1.1; Frost, 2011). Therefore, the relationship between higher mt substitution and diversification rates in frogs remains elusive, and unless a rampant extinction of basal neobatrachians accounts for the observed huge differences in diversity (number of species), it could be considered that substitution and diversification rates are decoupled in frogs. Unfortunately, the current fossil record is not broad enough to provide an answer to this question yet (Roček, 2000; Roček and Rage, 2000).

The studies here presented give compelling evidence of higher substitution rates in the mt genome of neobatrachian frogs, but nuclear genes did not show a clear trend, confirming previous observations (Hoegg et al., 2004). This might be the result of the different properties of mt and nuclear genomes, such as the recombination rate (virtually absent in mt DNA) or the effective population size (smaller in mt DNA), which can influence the effectiveness of selection upon these two genetic systems (Comeron et al., 2008).

An alternative explanation is that the substitution rate acceleration is general for both mt and nuclear genomes, but that these results are biased by the use of particular nuclear genes, which can hardly represent the entire nuclear genome and its complexity, with many genes obviously subjected to very disparate selective regimes (Arbiza et al., 2006). The possibility that the above results are slightly affected by some sort of phylogenetic artefact (Fuellen et al., 2001) cannot be totally ruled out, but the neobatrachian-specific higher mt substitution rates are reinforced by compelling evidence of relaxed purifying selection on mt proteins.

Furthermore, our results might be showing that selection could have relaxed also in nuclear genes, and thus justify the higher substitution rates found in the genes *rag2* and *slc8a1* in neobatrachians (Table 4.5). Data from additional nuclear genes, which are likely to be gathered soon in the context of new genome sequencing initiatives, hold the key to confirm or reject a putative general acceleration of evolutionary rates in neobatrachian frogs. Likewise, the clarification of the causes that relaxed purifying selection would need further, in-depth studies that investigate intrinsic and extrinsic factors that might have modified the fitness landscape of gene function (Lahti et al., 2009).

5.5. Perspectives in amphibian phylogenomics

No more than 20 years ago, molecular data produced by the Sanger method allowed a complete change in the way evolutionary relationships among amphibians were later approached (Hedges and Maxson, 1993). Since then, available molecular data is growing exponentially, and accordingly, phylogenetic studies use increasing amounts of data. However, most studies have typically used partial sequences from a few genes to addressed particular questions restricted to specific lineages (*e.g.,* García-París et al., 2003a; van der Meijden et al., 2005; Wiens et al., 2005); and relatively few studies have addressed higher-level phylogenetic relationships in amphibians. The study of deep phylogenetic relationships have sometimes relied on complete mt genome data, as they provide comparatively large number of molecular characters to estimate robust phylogenetic hypotheses, both among (*e.g*., Zardoya and Meyer, 2001; Zhang et al., 2005a) and within amphibian orders (*e.g.,* San Mauro et al., 2004b; Igawa et al., 2008; Zhang and Wake, 2009a, b).

In general, given limited time and resources, studies using Sanger sequencing suffer from a trade-off between taxon coverage and the number of characters per taxon that could be obtained. Both approaches, either sampling fewer characters for more species, or using mitogenome data for fewer taxa, have proven useful to answer phylogenetic questions at shallower and deeper levels of divergence, respectively. Furthermore, phylogenetic studies based on molecular data have contributed most to defining the amphibian tree of life as we currently view it, even though many questions still remain contentious.

While data from previous studies accumulated, continuously decreasing sequencing costs have allowed exponentially increasing the new sequence data for amphibians. As a consequence, the amount of sequence data in publicly available databases prompted large-scale phylogenetic studies using a broad taxon sampling with information from several loci. The first broad scale study on amphibian phylogeny was "The amphibian tree of life" (Frost et al., 2006), and included 552 species, with data from 152 morphological characters and two mt and five nuclear genes (up to 4.9 Kb per taxon; mean = 3.5 Kb). More recently, an updated version of a large-scale phylogeny for amphibians used 2,871 species, with data from 12 genes (mt and nuclear; up to 12.7 Kb per taxon; mean = 2.6 Kb) (Pyron and Wiens, 2011). These studies used all the available data in a supermatrix approach, trying to maximize both the number of taxa and the number of characters per taxon. Their contribution to amphibian systematics and higher-level taxonomy has been very important, for example, by pointing out to traditional non-monophyletic families and allowing them to be appropriately divided into smaller, monophyletic families.

Nevertheless, many of the questions that were unresolved in previous studies remained so in these large-scale studies, pointing out to the necessity of addressing contentious questions with studies designed on purpose. Here, complete mt genomes and nuclear loci have been used to address specific key questions in the phylogeny of frogs. This approach favoured having a large number of characters for few key taxa and a comprehensive matrix (missing data was minimal), as it provides a large number of characters to estimate robust phylogenetic hypotheses.

For many years, the amount of data to resolve a particular phylogenetic question has been a limiting factor, frequently producing contrasting hypotheses (Bininda-Emonds, 2011). However, the advent of new sequencing technologies and the completion of various genome projects is changing the field, and phylogenetics has moved a step forward, entering the era of genome-scale data sets: phylogenomics (Delsuc et al., 2005). Phylogenomics are a powerful tool to infer historical relationships and divergence times among species, but it will also permit studying the genetic basis of recent and rapid intraspecific adaptive changes, the genetic component of species' response to climate change and decline, or the phylogeography and population genetics, which are crucial to the assessment, monitoring, and management of biological diversity (Genome 10K Community of Scientists, 2009; Kumar et al., 2012). At the molecular level, phylogenomics can be used to identify traces of molecular adaptation, to infer evolutionary patterns of macromolecules, and predict gene functions (Kumar et al., 2012).

In the same way that genome sequences are excellent for evolutionary studies, the use of techniques of evolutionary analysis in comparative genomic studies is essential to understand issues of genome evolution, such as the origin and evolution of gene families, substitution patterns in non-coding DNA, etc. Furthermore, sequencing of complete nuclear genomes will permit obtaining information from non-coding DNA sequences (including regulatory regions, which

are known to play a fundamental role in evolution), or genes expressed at low levels (previously underrepresented in EST-based studies) (Genome 10K Community of Scientists, 2009; Telford and Copley, 2011). Both genome projects and transcriptome sequencing provide massive amounts of data for specific model organisms, but given the rapidly decreasing costs of next-generation sequencing, it is expected to be available for a wide range of taxa in the near future (Kumar et al., 2012).

The amount of genome-scale data available for amphibians is still restricted. Two genome projects (*Xenopus laevis* and *Silurana tropicalis*) and several EST collections are currently accessible in the GOLD Genome online database (Pagani et al., 2012), and the genome database of GenBank (Benson et al., 2010): EST collections include the salamanders *Ambystoma mexicanum* and *A. tigrinum,* and the anurans *Bufo marinus* (Bufonidae), *Hymenochirus curtipes* (Pipidae), *Rana pirica* (Ranidae), *Spea multiplicata* (Scaphiopodidae), and a *Xenopus laevis x muelleri* hybrid. An important step forward in the direction of massive sequencing of amphibian genomes was accomplished by the Genome 10K community of scientists (2009), who proposed to create a collection of tissue and DNA specimens for 10,000 vertebrate species, specifically designated for whole-genome sequencing. Their proposal included 1,760 species of amphibians (*ca.* 26% of the total species diversity), covering 301 genera and 50 families, including caecilians, salamanders and frogs (Genome 10K Community of Scientists, 2009). The project is ongoing, and currently 13 species are being sequenced (http://www.ldl.genomics.cn/page/pa-amphibian.jsp): the caecilian *Ichthyophis bannanicus* (Ichthyophiidae); the salamanders *Andrias davidianus* (Cryptobranchidae) and *Cynops orientalis* (Salamandridae); and the anurans *Ascaphus truei* (Leiopelmatidae), *Bombina orientalis* (Bombinatoridae), *Rhinophrynus dorsalis* (Pipidae), *Bufo gargarizans* (Bufonidae), *Atelopus zeteki* (Bufonidae), *Oophaga pumilio* (Dendrobatidae), *Gastrotheca cornuta* (Hemiphractidae), *Engystomops pustulosus* (Leiuperidae), *Eleutherodactylus coqui* (Eleutherodactylidae), and *Nanorana parkeri* (Dicroglossidae).

Genome-scale studies require a huge computational power for handling and analyzing the data, and despite the fact that the efficiency of both computers and software is continuously improving, computational power is currently one of the major limiting factors for phylogenomics (Bininda-Emonds, 2011). The millions of reads produced by next-generation sequencers need to be carefully analyzed and assembled, and in the case of complete genomes, assemblies need to be annotated; tasks that are not trivial and represent an active field of research (*e.g.,* Ruffalo et al., 2011).

Phylogenomic studies generally follow the supermatrix approach, where sequences for multiple genes are concatenated (Yang and Rannala, 2012). In the case of complete genomes, every locus is in principle available, with the exceptions of those that are specific to a lineage or have been secondarily lost. However, a previous step is needed to unambiguously discriminate between

orthologous and paralogous genes (Kumar et al., 2012). In the case of transcriptome data, not every locus is available, and typically, supermatrices possess very high proportion of missing data, with a large number of loci but low taxon coverage per locus (*e.g.,* Kocot et al., 2011; Smith et al., 2011). Even though in principle maximum likelihood and Bayesian inference methods could accommodate some sort of missing data as long as the matrix is largely informative, it is generally not well-understood how missing data impacts phylogenomic studies (Yang and Rannala, 2012).

On the other hand, phylogenomics are a very powerful tool, as they generally produce inferences that are very precise (variances are small) because stochastic error is greatly reduced, providing great statistical power to reject the null hypotheses (low p values) (Kumar et al., 2012). However, highly supported phylogenetic hypotheses have sometimes been reported depending on the evolutionary model and method of phylogenetic inference used (Jeffroy et al., 2006). This is produced by the fact that phylogenomic studies are more sensitive to model misspecifications, a phenomenon known as systematic error (Zhong et al., 2011; Kumar et al., 2012).

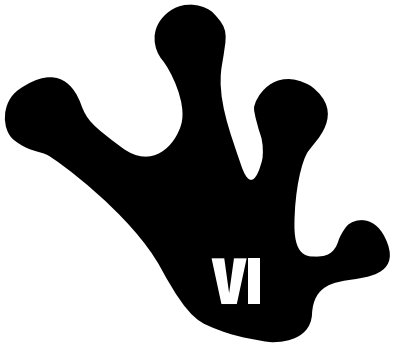
Systematic error is produced by departures of the real data from the assumptions of the evolutionary model, so it is common to every model-based phylogenetic method (Rodríguez-Ezpeleta et al., 2007). Even if the violations to the model are small in phylogenomics, the estimates become more and more biased with increasing sequence lengths, thus producing estimates that converge (with high support) to a biased value (Kumar et al., 2012). In other words, the variance of the estimate can be very low, but it does not mean that the estimate is any more accurate, as the mean value remains incorrect. Systematic error is introduced in phylogenetics simply because a model of evolution needs to be assumed; and although they can approximate to the truth, they are never so (Sullivan and Joyce, 2005; Kumar et al., 2012). For example, most evolutionary models assume that evolution can be modelled by a single, time-continuous Markov process, which is globally stationary, reversible, and homogeneous; implying that the evolutionary process cannot have been different neither across time nor across lineages (Jermiin et al., 2008). However, several studies have shown that this is unlikely to be so (*e.g.,* Phillips et al., 2004; Rodríguez-Ezpeleta et al., 2007; Dávalos and Perkins, 2008; Grievink et al., 2010; Nesnidal et al., 2010; Zhong et al., 2011). Methods to detect some model violations are starting to be developed (*e.g.,* Ho et al., 2006; Wang et al., 2011; Wu and Susko, 2011; Romiguier et al., 2012), but they are far from being comprehensive and widely used (Kumar et al., 2012).

A possible solution to reduce systematic error would be to use phylogenetic methods that are robust to model violations, even if they are less efficient (Yang and Rannala, 2012). On the other hand, some probabilistic models exist to relax at least some model assumptions (*e.g.,* Lartillot and Philippe, 2004; Blanquart and Lartillot, 2006; Jayaswal et al., 2007; Pagel and Meade, 2008; Jayaswal et al., 2011). These methods are highly parametric, but given a large dataset, the estimation of additional parameters should be generally well-tolerated (Telford and Copley, 2011).

However, they are computationally very demanding, and given the big size of the matrices, a trade-off must be made between size of the dataset and the sophistication of the method, using current software and hardware (Telford and Copley, 2011). Large matrices containing both long alignments and/ or a large number of taxa represent another challenge for computation, as it is well-known that the number of possible trees increases exponentially with the number of taxa (Bininda-Emonds, 2011).

Overall, phylogenomics represents a very promising field of research both for species-based studies, as well as for the development and refinement of methods of phylogenetic inference. In amphibians, genome-scale studies have already been used to address interesting phylogenetic questions on pipid relationships (Bewick et al., 2012), and genome evolution in salamanders (Sun et al., 2012).

Whole-genome sequencing is an interesting approach to tackle complex phylogenetic problems or long-standing controversial questions, but they are not suited (and seems at present pointless) to address taxonomically broad or population-based studies. Therefore, other strategies might be used to take advantage of the potential of the next-generation sequencing technology, such as PCR-based approaches that mix amplicons for several loci and specimens. Sequencing of multiplexed PCR products require some sort of *a posteriori* bioinformatic treatment to distinguish the reads from different specimens both by using labels and physical separation in the sequencing plate (Puritz et al., 2012), or by tags included in the primers of PCR amplifications (Bybee et al., 2011). These kind of taxonomic-broad approaches seem suitable to, for example, cover the vast diversity of neobatrachians (especially within Nobleobatrachia and Natatanura). At the same time, other studies using either complete genomes or transcriptomes would be very useful to tackle particularly difficult questions that would require sampling many genes, as would be the case of rapid radiations, where phylogenetic information for that particular event is likely to be scarce.



CONCLUSIONS

CONCLUSIONS

From the studies presented in this Ph. D. thesis concerning the phylogeny of frogs, the following conclusions can be drawn:

1. *Leiopelma* is the sister taxon of *Ascaphus* (forming the Amphicoela clade), and both together are the sister group of all other frogs, thus representing the most basal lineage of extant anurans.

2. The mt genome of *Leiopelma archeyi* displays a rearrangement affecting five genes upstream the control region, representing a derived gene order from that ancestral for vertebrates. This mt gene order is new for anurans, but evolved at least four times independently during vertebrate evolution. This evidence, along with comparative data for other vertebrates suggests that the 5' end of the control region is a hot spot for gene order change.

3. Living frogs form five major lineages, which branch successively as (i) Amphicoela, (ii) Discoglossoidea, (iii) Pipoidea, (iv) Pelobatoidea, and (v) Neobatrachia.

4. Within Pipoidea, Rhinophrynidae is the sister family of Pipidae. Within this latter family, species showing a current South American distribution (genus *Pipa*) are the sister group of an African lineage, in which dactylethrines (*Xenopus* + *Silurana*) are sister to hymenochirines (*Hymenochirus* + *Pseudhymenochirus*).

5. In contrast to all other pipids that possess a highly specialized sound production mechanism independent of air movement, sound production in *Pseudhymenochirus merlini* is air-driven. This represents a reversal to the ancestral condition that occurs in all other non-pipid frogs: However, anatomical studies clearly reveal the pipid nature of the *Pseudhymenochirus* larynx. Therefore, sound production in *Pseudhymenochirus* represents an evolutionary innovation that evolved constrained by its phyletic heritage, under the restrictions imposed by the larynx of pipids.

6. Within Neobatrachia, *Heleophryne* is the sister group of all other neobatrachian frogs. The clade formed by *Calyptocephalella* (Calyptocephalellidae) and *Lechriodus* (Limnodynastidae) is the sister group of Nobleobatrachia. Nobleobatrachia is a highly speciose clade, in which *Telmatobius* (Ceratophryidae) and *Duttaphrynus* (Bufonidae) are closely related to the exclusion of *Hyla* (Hylidae). The phylogenetic position of the family Sooglossidae could not be unambiguously established, but it is recovered as the sister group of Ranoides (a second speciose clade), where Microhyloidea (*Microhyla* + *Kaloula*) is the sister group of the clade Natatanura, containing (Dicroglossidae + (Ranidae + (Mantellidae + Rhacophoridae))).

7. All new mt gene rearrangements reported are consistent with the tandem duplication–random loss model, which is regarded as the main mechanisms of gene order change in vertebrate mitochondria.

8. The *LTPF* tRNA gene cluster typical of neobatrachian mt genomes was already present in *Heleophryne*, thus suggesting that it represents a molecular synapomorphy for the clade. Nevertheless the mt genomes of *Heleophryne regis* and *Lechriodus melanopyga* depart from the ancestral gene order of neobatrachians by rearrangements that involve the *IQM* and *WANCY* tRNA clusters, and the origin of replication of the light strand, which has been shown to be a hot spot for gene order change.

9. Neobatrachian frogs exhibit a significantly higher mt substitution rates compared to non-neobatrachians, and they became accelerated in the origin of the group (in the Early Jurassic). The cause that triggered such rate acceleration is unknown, but the relaxation of purifying selection acting on mt protein-coding genes in the origin of the group could (at least in part) account the observed pattern of substitution rates.

10. Both complete mt genomes and the nine nuclear protein-coding genes demonstrate good phylogenetic performance and overall congruence, demonstrating that they are appropriate markers to estimate deep phylogenetic relationships among frogs.

CONCLUSIONES

De los estudios presentados en esta tesis doctoral acerca de la filogenia de anuros, pueden extraerse las siguientes conclusiones:

1. *Leiopelma* es el taxón hermano de *Ascaphus* (formando el clado Amphicoela), y ambos son el grupo hermano del resto de los anuros, de modo que representan el linaje más basal entre los anuros actualmente existentes.

2. El genoma mt de *Leiopelma archeyi* presenta un reordenamiento génico que afecta a cinco genes que se encuentran al lado de la región control, y representa un orden génico derivado respecto al ancestral de vertebrados. Este orden génico mt es nuevo para anuros, pero evolucionó al menos cuatro veces independientemente durante la evolución de los vertebrados. Esta evidencia, junto con datos comparativos del resto de vertebrados, sugiere que la región 5' de la región control es un punto caliente para la reordenación génica.

3. Los anuros actuales forman cinco linajes principales, que se ramifican sucesivamente como (i) Amphicoela, (ii) Discoglossoidea, (iii) Pipoidea, (iv) Pelobatoidea, y (v) Neobatrachia.

4. Dentro de Pipoidea, Rhinophrynidae es la familia hermana de Pipidae. Dentro de esta última familia, las especies de distribución actual sudamericana (género *Pipa*) son el grupo hermano del linaje africano, en el que los dactylethrines (*Xenopus* + *Silurana*) son grupo hermano de los hymenochirines (*Hymenochirus* + *Pseudhymenochirus*).

5. Al contrario de lo que ocurre en el resto de pípidos que cuentan con un mecanismo de producción del sonido altamente especializado e independiente del movimiento de aire, la producción sonora en *Pseudhymenochirus merlini* ocurre mediante un flujo de aire. Esto representa una reversión a la condición ancestral que encontramos en el resto de las ranas no pertenecientes a la familia Pipidae. Sin embargo, los estudios anatómicos dejan clara la naturaleza de tipo pípido de la laringe en *Pseudhymenochirus*. Por lo tanto, la producción sonora en *Pseudhymenochirus* representa una innovación evolutiva que evolucionó constreñida bajo su herencia filética, bajo unas restricciones impuestas por la laringe de los pípidos.

6. Dentro de Neobatrachia, *Heleophryne* es el grupo hermano de resto de neobatráceos. El clado formado por *Calyptocephalella* (Calyptocephalellidae) y *Lechriodus* (Limnodynastidae) es el grupo hermano de Nobleobatrachia. Nobleobatrachia es un clado muy diversificado, donde *Telmatobius* (Ceratophryidae) y *Duttaphrynus* (Bufonidae) están más próximamente relacionados,

excluyendo a *Hyla* (Hylidae). La posición filogenética de la familia Sooglossidae no pudo ser establecida inequívocamente, pero aparece como el grupo hermano de Ranoides (un segundo clado altamente diversificado), donde Microhyloidea (*Microhyla* + *Kaloula*) es el grupo hermano del clado Natatanura, que contiene (Dicroglossidae + (Ranidae + (Mantellidae + Rhacophoridae))).

7. Todos los nuevos órdenes génicos descubiertos son consistentes con el modelo de duplicación en tándem y pérdida aleatoria, que es considerado el principal mecanismo de reordenamiento génico en la mitocondria de vertebrados.

8. El grupo de genes de tRNA *LTPF*, típico del genoma mt de neobatráceos, estaba ya presente en *Heleophryne*, sugiriendo que se trata de una sinapomorfía molecular para el grupo. A pesar de ello, los genomas de *Heleophryne regis* y *Lechriodus melanopyga* difieren del orden ancestral de los neobatráceos en reordenamientos que afectan a los grupos de genes de tRNA *IQM* y *WANCY*, así como al origen de replicación de la cadena ligera, habiéndose demostrado que éste es un punto caliente para la reordenación génica.

9. Los anuros neobatráceos presentan una tasa de sustitución mitocondrial significativamente mayor respecto a los no-neobatráceos, que se aceleró en el origen del grupo (en el Jurásico temprano). El desencadenante de tal aceleración se desconoce, pero la relajación de la selección purificadora que actúa sobre los genes mt codificantes para proteínas podría explicar, al menos parcialmente, el patrón observado de tasas de sustitución.

10. Tanto los genomas mt completos, como los nueve genes nucleares codificantes para proteínas tienen un buen rendimiento filogenético, y son congruentes en su conjunto, demostrando que representan marcadores apropiados para estimar relaciones filogenéticas profundas entre los anuros.

REFERENCES

REFERENCES

Abascal F, Zardoya R, and Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104-2105.

Abascal F, Posada D, Knight RD, and Zardoya R. 2006. Parallel evolution of the genetic code in arthropod mitochondrial genomes. PLoS Biol 4:e127.

Abascal F, Zardoya R, and Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nuc Acids Res 38:W7-W13.

Adachi J, and Hasegawa M. 1995. Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: heterogeneity among amino acid sites. J Mol Evol 40:622-628.

Adachi J, and Hasegawa M. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J Mol Evol 42:459-468.

Agrawal A, Eastman QM, and Schatz DG. 1998. Implications of transposition mediated by V(D) J-recombination proteins RAG1 and RAG2 for origins of antigen-specific immunity. Nature 394:744-751.

Akaike H. 1973. Information theory as an extension of the maximum likelihood principle. In Petrov BN, and Csaki F, editors. Second international symposium of information theory. Akademiai Kiado, Budapest.

Alam MS, Kurabayashi A, Hayashi Y, Sano N, Khan MMR, Fujii T, and Sumida M. 2010. Complete mitochondrial genomes and novel gene rearrangements in two dicroglossid frogs, *Hoplobatrachus tigerinus* and *Euphlyctis hexadactylus*, from Bangladesh. Genes Genet Syst 85:219-232.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, and Walter P. 2002. Molecular biology of the cell. Garland Science Cop., New York.

Alfaro ME, Zoller S, and Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol Biol Evol 20:255-256.

Allen JF. 2003. The function of genomes in bioenergetic organelles. Philos Trans R Soc B 358:19-38.

Altig RI, and Johnston GF. 1989. Guilds of anuran larvae: Relationships among developmental modes, morphologies, and habitats. Herpetol Monogr 3:81-109.

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-410.

Álvarez I, and Wendel JF. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogenet Evol 29:417-434.

Amer SAM, and Kumazawa Y. 2007. The mitochondrial genome of the lizard *Calotes versicolor* and a novel gene inversion in South Asian draconine agamids. Mol Biol Evol 24:1330-1339.

Amphibiaweb. 2011. Information on amphibian biology and conservation [web application]. Berkeley, California: Amphibiaweb. Available: http://amphibiaweb.org/.

Anderson JS, Reisz RR, Scott D, Frobisch NB, and Sumida SS. 2008. A stem batrachian from the Early Permian of Texas and the origin of frogs and salamanders. Nature 453:515-518.

Anisimova M, Bielawski JP, and Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585-1592.

Anisimova M, and Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol 55:539-552.

Arbiza L, Dopazo J, and Dopazo H. 2006. Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. PLoS Comput Biol 2:e38.

Archey G. 1922. The habitat and life history of *Leiopelma hochstetteri*. Rec Canterbury Mus 2:59-71.

Báez AM. 2000. Tertiary anurans from South America. Pp. 1388-1401. In Heatwole H, and Carroll RL, editors. Amphibian biology. Surrey Beatty, Chipping Norton, Australia.

Báez AM, Trueb L, and Calvo JO. 2000. The earliest known pipoid frog from South America: A new genus from the middle Creataceous of Argentina. J Vertebr Paleontol 20:490-500.

Báez AM, and Harrison T. 2005. A new pipine frog from an eocene crater lake in North-central Tanzania. Palaeontology 48:723-737.

Barbadillo LJ, García-París M, and Sanchíz B. 1997. Orígenes y relaciones evolutivas de la herpetofauna ibérica. Pp. 47-100. In Pleguezuelos JM, editor. Distribución y biogeografía de los anfibios y reptiles en España y Portugal. Universidad de Granada, Granada.

Barraclough TG, and Savolainen V. 2001. Evolutionary rates and species diversity in flowering plants. Evolution 55:677-683.

Batni S, Scalzetti L, Moody SA, and Knox BE. 1996. Characterization of the *Xenopus* rhodopsin gene. J Biol Chem 271:3179-3186.

Beatty J, and Fink WL. 1979. Review of "simplicity", by Elliot Sober. Syst Zool 28:643-651.

Bell BD. 1978. Observations in the ecology and reproduction of the New Zealand leiopelmatid frogs. Herpetologica 34:340-454.

Bell BD, and Wassersug RJ. 2003. Anatomical features of *Leiopelma* embryos and larvae: implications for anuran evolution. J Morphol 256:160-170.

Bensasson D, Zhang DX, Hartl DL, and Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol 16:314–321.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, and Sayers EW. 2010. GenBank. Nuc Acids Res 39:D32-D37.

Benton MJ. 1990. Phylogeny of the major tetrapod groups: morphological data and divergence dates. J Mol Evol 30:409-424.

Benton MJ, and Donoghue PCJ. 2007. Paleontological evidence to date the tree of life. Mol Biol Evol 24:26-53.

Berry V, and Gascuel O. 1996. On the interpretation of bootstrap trees: appropriate threshold of clade selection and induced gain. Mol Biol Evol 13:999-1011.

Bewick AJ, Chain FJJ, Heled J, and Evans BJ. 2012. The pipid root. Syst Biol in press.

Bidnenko V, Ehrlich SD, and Michel B. 2002. Replication fork collapse at replication terminator sequences. EMBO J 21:3898-3907.

Biju SD, and Bossuyt F. 2003. New frog family from India reveals an ancient biogeographical link with the Seychelles. Nature 425:711-714.

Bininda-Emonds O. 2011. Inferring the tree of life: chopping a phylogenomic problem down to size? BMC Biology 9:59.

Blackburn DC, Bickford DP, Diesmos AC, Iskandar DT, and Brown RM. 2010. An ancient origin for the enigmatic flat-headed frogs (Bombinatoridae: *Barbourula*) from the islands of Southeast Asia. PLoS ONE 5:e12090.

Blanquart S, and Lartillot N. 2006. A bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. Mol Biol Evol 23:2058-2071.

Blaustein AR, and Wake DB. 1990. Declining amphibian populations – a global phenomenon. Trends Ecol Evol 5:203-204.

Blommers-Schlösser RMA. 1993. Systematic relationships of the Mantellinae Laurent 1946 (Anura Ranoidea). Ethol Ecol Evol 5:199-218.

Bogenhagen DF. 1999. Repair of mtDNA in vertebrates. Am J Hum Genet 64:1276-1281.

Boore JL, and Brown WM. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr Opin Genet Dev 8:668-674.

Boore JL. 1999. Animal mitochondrial genomes. Nuc Acids Res 27:1767-1780.

Boore JL. 2000. The duplication/ random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. Pp. 133-147. In Sankoff D, and Nadeau JH, editors. Comparative genomics. Kluwer Academic Publisher, Dordrecht.

Boore JL, Macey RJ, and Medina M. 2005. Sequencing and comparing whole mitochondrial genomes of animals. Method Enzymol 395:311-348.

Bossuyt F, and Roelants K. 2009. Frogs and toads. Pp. 357-364. In Hedges SB, and Kumar S, editors. The Timetree of Life. Oxford University Press, New York.

Boulenguer GA. 1882. Catalogue of the Batrachia Salientia s. Ecaudata in the collection of the British Museum. Taylor and Francis, London.

Bowes JB, Snyder KA, Segerdell E, Jarabek CJ, Azam K, Zorn AM, and Vize PD. 2010. Xenbase: gene expression and improved integration. Nuc Acids Res 38:D607-612.

Bowmaker M, Yang MY, Yasukawa T, Reyes A, Jacobs HT, Huberman JA, and Holt IJ. 2003. Mammalian mitochondrial DNA replicates bidirectionally from an initiation zone. J Biol Chem 278:50961-50969.

Breton S, Beaupré HD, Stewart DT, Hoeh WR, and Blier PU. 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends Genet 23:465-474.

Bromham L, Rambaut A, and Harvey P. 1996. Determinants of rate variation in mammalian DNA sequence evolution. J Mol Evol 43:610-621.

Bromham L, and Leys R. 2005. Sociality and the rate of molecular evolution. Mol Biol Evol 22:1393-1402.

Bromham L. 2009a. Why do species vary in their rate of molecular evolution? Biol Lett 5:401-404.

Bromham L. 2009b. Putting the 'bio' into bioinformatics. Biol Lett 5:391-393.

Brown JM, Hedtke SM, Lemmon AR, and Lemmon EM. 2010. When trees grow too long: investigating the causes of highly inaccurate Bayesian branch-length estimates. Syst Biol 59:145-161.

Brown WM, George MJ, and Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76:1967-1971.

Brown WM. 1983. Evolution of animal mitochondrial DNA. Pp. 62-88. In Nei M, and Koehn RK, editors. Evolution of genes and proteins. Sinauer Associates, Sunderland, Massachusetss.

Bruford MW, Hanotte O, and Brookfield JFY. 1992. Single-locus and multilocus DNA fingerprint. In Hoelzel AR, editor. Molecular genetic analysis of populations: a practical approach. IRL Press, Oxford.

Buckley TR, and Cunningham CW. 2002. The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. Mol Biol Evol 19:394-405.

Bybee SM, Bracken-Grissom HD, Hermansen RA, Clement MJ, Crandall KA, and Felder DL. 2011. Directed next generation sequencing for phylogenetics: an example using Decapoda (Crustacea). Zool Anz 250:497-506.

Cameron SL, Lambkin CL, Barker SC, and Whiting MF. 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Syst Entomol 32:40-59.

Camin JH, and Sokal RR. 1965. A method for deducing branching sequences in phylogeny. Evolution 19:311-326.

Cannatella DC. 1985. A phylogeny of primitive frogs (Archaeobatrachia). PhD dissertation. Department of Systematics and Ecology. University of Kansas, Lawrence.

Cannatella DC, and Trueb L. 1988a. Evolution of pipoid frogs: Intergeneric relationships of the aquatic frog family Pipidae (Anura). Zool J Linn Soc 94:1-38.

Cannatella DC, and Trueb L. 1988b. Evolution of pipoid frogs: Morphology and phylogenetic relationships of *Pseudhymenochirus*. J Herpetol 22:439-456.

Cannatella DC, and de Sá RO. 1993. *Xenopus laevis* as a model organism. Syst Biol 42:476-507.

Cao Y, Adachi J, Janke A, Pääbo S, and Hasegawa M. 1994. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. J Mol Evol 39:519-527.

Carey C, and Alexander MA. 2003. Climate change and amphibian declines: is there a link? Divers Distrib 9:111-121.

Carroll RL. 2000. *Eocaecilia* and the origin of caecilians. Pp. 1402-14011. In Heatwole H, and Carroll RL, editors. Amphibian biology. Surrey Beatty, Chipping Norton, Australia.

Carroll RL. 2001. The origin and early radiation of terrestrial vertebrates. J Paleontol 75:1202-1213.

Carroll RL, Boisvert C, Bolt J, Green DM, Philp N, Rolian C, Schoch RR, and Tarenko A. 2004. Changing patterns of ontogeny from osteolepiform fish through Permian tetrapods as a guide to the early evolution of land vertebrates. Pp. 321-343. In Arratia G, Wilson MVH, and Cloutier R, editors. Recent advances in the origin and early radiation of vertebrates, Pfeil, München.

Carroll RL. 2007. The Palaeozoic ancestry of salamanders, frogs and caecilians. Zool J Linn Soc 150:1-140.

Carroll RL. 2009. The rise of amphibians: 365 million years of evolution. The Johns Hopkins University Press, Baltimore.

Castellana S, Vicario S, and Saccone C. 2011. Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein coding genes. Genome Biol Evol 3:1067-1079.

Castoe TA, Jiang ZJ, Gu W, Wang ZO, and Pollock DD. 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. PLoS ONE 3:e2201.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-552.

Cavalli-Sforza LL, and Edwards AWF. 1967. Phylogenetic analysis: models and estimation producedures. Evolution 21:550-570.

Chabanaud P. 1921. Contribution a l'étude de la faune herpétologique de l'Afrique occientale. Bull Com Études Hist Scient Afr Occid Franç 1921:445-472.

Chang BSW, and Campbell DL. 2000. Bias in phylogenetic reconstruction of vertebrate rhodopsin sequences. Mol Biol Evol 17:1220-1231.

Chiari Y, Vences M, Vieites DR, Rabemananjara F, Bora P, Ramilijaona Ravoahangimalala O, and Meyer A. 2004. New evidence for parallel evolution of colour patterns in Malagasy poison frogs (*Mantella*). Mol Ecol 13:3763-3774.

Clayton DA. 1982. Replication of animal mitochondrial DNA. Cell 28:693-705.

Clayton DA. 1984. Transcription of the mammalian mitochondrial genome. Annu Rev Biochem 53:573-594.

Clegg MT, Gaut BS, Learn GH, and Morton BR. 1994. Rates and patterns of chloroplast DNA evolution. Proc Natl Acad Sci USA 91:6795-6801.

Colgan DJ, Ponder WF, and Eggler PE. 2000. Gastropod evolutionary rates and phylogenetic relationships assessed using partial 28S rDNA and histone H3 sequences. Zool Scr 29:29-63.

Collura RV, and Stewart C-B. 1995. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominids. Nature 378:485–489.

Comeron JM, Williford A, and Kliman RM. 2008. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage. Heredity 100:19-31.

Cope ED. 1865. Sketch of primary groups of Batrachia Salientia. Nat History Review 5:97-120.

Cotton JA, and Page RDM. 2002. Going nuclear: gene family evolution and vertebrate phylogeny reconciled. Proc R Soc Lond B 269:1555-1561.

Cummings MP, and Meyer A. 2005. Magic bullets and golden rules: data sampling in molecular phylogenetics. Zoology 108:329-336.

Curole JP, and Kocher TD. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol Evol 14:394-398.

D'Onorio de Meo P, D'Antonio M, Griggio F, Lupi R, Borsani M, Pavesi G, Castrignanò T, Pesole G, and Gissi C. 2012. MitoZoa 2.0: a database resource and search tools for comparative and evolutionary analyses of mitochondrial genomes in Metazoa. Nuc Acids Res 40:D1168-D1172.

da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, and Antunes A. 2008. The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics 9:119.

Darst CR, and Cannatella DC. 2004. Novel relationships among hyloid frogs inferred from 12S and 16S mitochondrial DNA sequences. Mol Phylogenet Evol 31:462-475.

Darwin C. 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. John Murray, London.

Daszak P, Cunningham AA, and Hyatt AD. 2003. Infectious disease and amphibian population declines. Divers Distrib 9:141-150.

Dávalos LM, and Perkins SL. 2008. Saturation and base composition bias explain phylogenomic conflict in *Plasmodium*. Genomics 91:433-442.

Davies TJ, Savolainen V, Chase MW, Moat J, and Barraclough TG. 2004. Environmental energy and evolutionary rates in flowering plants. Proc R Soc Lond B 271:2195-2200.

de Grey ADNJ. 2005. Forces maintaining organellar genomes: is any as strong as genetic code disparity or hydrophobicity? BioEssays 27:436-446.

de Queiroz K, and Poe S. 2001. Philosophy and phylogenetic inference: A comparison of likelihood and parsimony methods in the context of Karl Popper's writings on corroboration. Syst Biol 50:305-321.

de Sá RO, and Hillis DM. 1990. Phylogenetic relationships of the pipid frogs *Xenopus* and *Silurana*: an integration of ribosomal DNA and morphology. Mol Biol Evol 7:365-376.

de Sá RO, and Swart CC. 1999. Development of the suprarostral plate of pipoid frogs. J Morphol 240:143–153.

Deen PMT, Terwel D, Bussemakers MJM, Roubos EW, and Martens GJM. 1991. Structural analysis of the entire proopiomelanocortin gene of *Xenopus laevis*. Eur J Biochem 201:129-137.

Delsuc F, Brinkmann H, and Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 6:361-375.

Denver DR, Morris K, Lynch M, Vassilieva LL, and Thomas W. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabidits elegans*. Science 289:2342-2344.

Denver DR, Morris K, Lynch M, and Thomas W. 2004. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature 430:679-682.

Doda JN, Wright CT, and Clayton DA. 1981. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc Natl Acad Sci USA 78:6116-6120.

Donoghue PCJ, and Benton MJ. 2007. Rocks and clocks: calibrating the tree of life using fossils and molecules. Trends Ecol Evol 22:424-431.

Doucet-Beaupré H, Breton S, Chapman EG, Blier PU, Bogan AE, Stewart DT, and Hoeh WR. 2010. Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. BMC Evol Biol 10:50-69.

Dowton M, and Austin AD. 1999. Evolutionary dynamics of a mitochondrial rearrangement "hot spot" in the Hymenoptera. Mol Biol Evol 16:298-309.

Dowton M, and Campbell NJH. 2001. Intramitochondrial recombination: is it why some mitochondrial genes sleep around? Trends Ecol Evol 16.

Driskell AC, Ané C, Burleigh JG, McMahon MM, O'Meara BC, and Sanderson MJ. 2004. Prospects for building the tree of life from large sequence databases. Science 306:1172-1174.

Drummond AJ, Ho SYW, Phillips MJ, and Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol 4:e88.

Drummond AJ, and Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214-721.

Dubois A. 1985. Miscellanea nomenclatorica batrachologica (VII). Alytes 4:61-78.

Dubois A. 2004. Developmental pathway, speciation and supraspecific taxonomy in amphibians 1. Why are there so many frog species in Sri Lanka? Alytes 22:19-37.

Duellman WE. 1975. On the classification of frogs. Occ Pap Mus Nat Hist Univ Kansas 42:1-14.

Duellman WE, and Maness SJ. 1980. The reproductive behavior of some hylid marsupial frogs. J Herpetol 14:213-222.

Duellman WE, and Trueb L. 1986. Biology of amphibians. MacGraw-Hill, New York.

Dunn ER. 1948. American frogs of the family Pipidae. Am Mus Novit 1384:1-13.

Dutta SK, Vasudevan K, Chaitra MS, Shanker K, and Aggrwal RK. 2004. Jurassic frogs and the evoution of amphibian endemism in the Western Ghats. Curr Sci 86:211-216.

Eck RV, and Dayhoff MO. 1966. Atlas of protein sequence and structure 1966. National Biomedical Research Foundation, Silver Spring, Maryland.

Eddy S, and Durbin R. 1994. RNA sequence analysis using covariance models. Nuc Acids Res 22:2079-2088.

Edwards AWF, and Cavalli-Sforza LL. 1963. The reconstruction of evolution. Ann Hum Genet 27:105-106.

Edwards AWF, and Cavalli-Sforza LL. 1964. Reconstruction of evolutionary trees. In Heywood VH, and MCNeill J, editors. Phenetic and Phylogenetic Classification. Systematics Association, London.

Edwards SV, Liu L, and Pearl DK. 2007. High-resolution species trees without concatenation. Proc Natl Acad Sci USA 104:5936–5941.

Efron B. 1979. Bootstrap methods: another look at the jackknife. Ann Stat 7:1-26.

Emerson SB. 1984. Morphological variation in frog pectoral girdles: testing alternatives to a traditional adaptive explanation. Evolution 38:376-388.

Emerson SB. 1988. Convergence and morphological constraint in frogs: Variation in postcranial morphology. Field Zool 43:1-19.

Eo SH, and DeWoody JA. 2010. Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. Proc R Soc Lond B 277:3587-3592.

Erixon P, Svennblad B, Britton T, and Oxelman B. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst Biol 52:665-673.

Estes R, and Reig OA. 1973. The early fossil record of frogs: a review of the evidence. Pp. 11-63. In Vial JL, editor. Evolutionary biology of the anurans: contemporary research on major problems. University of Missouri Press, Columbia.

Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, and Cannatella DC. 2004. A mitochondrial DNA phylogeny of African clawed frogs: Phylogeography and implications for polyploid evolution. Mol Phylogenet Evol 33:197-213.

Evans BJ, Kelley DB, Melnick DJ, and Cannatella DC. 2005a. Evolution of *rag-1* in polyploid clawed frogs. Mol Biol Evol 22:1193-1207.

Evans BJ. 2008. Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). Front Biosci 13:4687-4706.

Evans SE, Milner AR, and Mussett F. 1990. A discoglossid frog from the Middle Jurassic of England. Palaeontology 33:299-311.

Evans SE, Milner AR, and Werner C. 1996. Sirenid salamanders and a gymnophionan amphibian from the Cretaceous of the Sudan. Palaeontology 39:77-95.

Evans SE, and Borsuk-Bialynicka M. 1998. A stem-group frog from the early Triassic of Poland. Acta Palaeontol Pol 43:573-580.

Evans SE, Lally C, Chure DC, Elder A, and Maisano JA. 2005b. A Late Jurassic salamander (Amphibia: Caudata) from the Morrison formation of North America. Zool J Linn Soc 143:599-616.

Faivovich J, Haddad CFB, Garcia PCA, Frost DR, Campbell JA, and Wheeler WC. 2005. Systematic review of the frog family Hylidae, with special reference to Hylinae: a phylogenetic analysis and taxonomic revision. Bull Amer Mus Nat Hist 294:1–240.

Farris JS. 1970. Methods for computing Wagner trees. Syst Zool 19:83-92.

Farris JS. 1977. Phylogenetic analysis under Dollo's law. Syst Zool 26:77-88.

Farris JS. 1983. The logical basis of phylogenetic analysis. Pp. 7-36. In Platnick NI, and Funk VA, editors. Advances in cladistics 2. Columbia University Press, New York.

Feller AE, and Hedges SB. 1998. Molecular evidence for the early history of living Amphibians. Mol Phylogenet Evol 9:509–516.

Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78:737-756.

Felsenstein J. 1978a. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401-410.

Felsenstein J. 1978b. The number of evolutionary trees. Syst Zool 27:27-33.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368-376.

Felsenstein J. 1985a. Phylogenies and the comparative method. Am Nat 125:1-15.

Felsenstein J. 1985b. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

Felsenstein J, and Kishino H. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. Syst Biol 42:193-200.

Felsenstein J. 2004. Inferring phylogenies. Sinauer Associates Inc., Sunderland, Massachusetts.

Fenn JD, Song H, Cameron SL, and Whiting MF. 2008. A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. Mol Phylogenet Evol 49:59-68.

Fernández-Silva P, Enriquez JA, and Montoya J. 2003. Replication and transcription of mammalian mitochondrial DNA. Exp Physiol 88:41-56.

Fisher RA. 1922. On the mathematical foundations of theoretical statistics. Phil Trans R Soc A 222: 309-368.

Fisher-Reid MC, and Wiens J. 2011. What are the consequences of combining nuclear and mitochondrial data for phylogenetic analysis? Lessons from *Plethodon* salamanders and 13 other vertebrate clades. BMC Evol Biol 11:300.

Fitch WM, and Margoliash E. 1967. A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome *c* as a model case. Biochem Genet 1:65-71.

Fitch WM. 1970. Distinguishing homologous from analogous proteins. Syst Zool 19:99-113.

Fitch WM. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst Zool 20:406-416.

Fonseca MM, 2011. Mitochondrial DNA evolution: replication and gene overlapping. PhD dissertation. University of Porto, Porto.

Ford LS. 1989. The phylogenetic position of poison-dart frogs (Dendrobatidae): reassessment of the neobatrachian phylogeny with commentary on complex character systems. PhD dissertation. The University of Kansas, Lawrence.

Ford LS, and Cannatella DC. 1993. The major clades of frogs. Herpetol Monogr 7:93-117.

Foster PG, Jermiin LS, and Hickey DA. 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J Mol Biol 44:282-288.

Frost DR, Grant T, Faivovich J, Bain RH, Haas A, Haddad CFB, de Sá RO, Channing A, Wilkinson M, Donnellan SC, Raxworthy CJ, Campbell JA, Blotto BL, Moler P, Drewes RC, Nussbaum R, Lynch JD, Green DM, and Wheeler WC. 2006. The amphibian tree of life. Bull Amer Mus Nat Hist 297:1-370.

Frost DR. 2011. Amphibian species of the world: an online reference. American Museum of Natural History, New York, USA.

Fuellen G, Wägele J-W, and Giegerich R. 2001. Minimum conflict: a divide-and-conquer approach to phylogeny estimation. Bioinformatics 17:1168-1178.

Funk DJ, and Omland KE. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol Syst 34:397-423.

Gabaldón T. 2008. Large-scale assignment of orthology: back to phylogenetics? Genome Biol 9:235.

Gach MH, and Brown WM. 1997. Characteristics and distribution of large tandem duplications in brook stickleback (*Culea inconstans*) mitochondrial DNA. Genetics 145:383-394.

Gaffney ES. 1979. An introduction to the logic of phylogeny reconstruction. Pp. 79-111. In Cracraft J, and Eldredge N, editors. Phylogenetic analysis and paleontology. Columbia University Press, New York.

Gao KQ, and Wang Y. 2001. Mesozoic anurans from Liaoning province, China, and phylogenetic relationships of archaeobatrachian anuran clades. J Vert Paleontol 21:460-476.

Gao KQ, and Shubin NH. 2003. Earliest known crown-group salamanders. Nature 422:424-428.

García-París M, Alcobendas M, Buckley D, and Wake DB. 2003a. Dispersal of viviparity across contact zones in iberian populations of fire salamanders (*Salamandra*) inferred from discordance of genetic and morphological traits. Evolution 57:129-143.

García-París M, Buchholz DR, and Parra-Olea G. 2003b. Phylogenetic relationships of Pelobatoidea re-examined using mtDNA. Mol Phylogenet Evol 28:12-23.

Gatesy J, Matthee C, DeSalle R, and Hayashi C. 2002. Resolution of a supertree/supermatrix paradox. Syst Biol 51:652–664.

Gatesy J, and Baker RH. 2005. Hidden likelihood support in genomic data: can fortyfive wrongs make a right? Syst Biol 54:483–492.

Gaut BS, Muse SV, Clark WD, and Clegg MT. 1992. Relative rates of nucleotide substitution at the *rbc*l locus of monocotyledonous plants. J Mol Evol 35:292-303.

Gaut BS, and Lewis PO. 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. Mol Biol Evol 12:152-162.

Genome 10K Community of Scientists. 2009. Genome 10K: a proposal to obtain whole-genome sequence for 10 000 vertebrate species. J Hered 100:659-674.

Gillespie JH. 1986. Rates of molecular evolution. Annu Rev Ecol Syst 17:637-665.

Gissi C, San Mauro D, Pesole G, and Zardoya R. 2006. Mitochondrial phylogeny of Anura (Amphibia): A case study of congruent phylogenetic reconstruction using amino acid and nucleotide characters. Gene 366:228-237.

Gissi C, Iannelli F, Pesole G. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101:301-20.

Glaw F, and Vences M. 1991. Bioacoustic differentiation in painted frogs (*Discoglossus*). Amphibia-Reptilia 12:385-394.

Golding GB. 1983. Estimates of DNA and protein sequence divergence: an examination of some assumptions. Mol Biol Evol 1:125-142.

Goldman N. 1993. Statistical tests of models of DNA substitution. J Mol Evol 36:182-198.

Goldman N. 1998. Phylogenetic information and experimental design in molecular systematics. Proc R Soc Lond B 265:1779-1786.

Goldman N, Anderson JP, and Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. Syst Biol 49:652-670.

Götz R, Raulf F, and Schartl M. 1992. Brain-derived neurotrophic factor is more highly conserved in structure and function than nerve growth factor during vertebrate evolution. J Neurochem 59:432-442.

Gould SJ. 1980. The panda's thumb. W. W. Norton & Company, New York.

Grant T, Frost DR, Caldwell JP, Gagliardo R, Haddad CFB, Kok PJR, Means DB, Noonan BP, Schargel WE, and Wheeler WC. 2006. Phylogenetic systematics of dart-poison frogs and their relatives (Amphibia: Athesphatanura: Dendrobatidae). Bull Amer Mus Nat Hist 299:1-262.

Graur D, and Martin W. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet 20:80-86.

Graybeal A. 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. Syst Biol 43:174-193.

Graybeal A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst Biol 47:9-17.

Green DM, Sharbel TF, Hitchmough RA, and Daugherty CH. 1989. Genetic variation in the genus *Leiopelma* and relationships to other primitive frogs. J Zoolog Syst Evol Res 27:65-79.

Green DM, and Cannatella DC. 1993. Phylogenetic significance of the amphicoelous frogs, Ascaphidae and Leiopelmatidae. Ethol Ecol Evol 5:233-245.

Grievink LS, Penny D, Hendy MD, and Holland BR. 2010. Phylogenetic tree reconstruction accuracy and model fit when proportions of variable sites change across the tree. Syst Biol 59:288-297.

Groth JG, and Barrowclough GF. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol Phylogenet Evol 12:115-123.

Gyllensten U, Wharton D, Josefsson A, and Wilson AC. 1991. Paternal inheritance of mitochondrial DNA in mice. Nature 352:255-257.

Haas A. 1997. The larval hyobranchial apparatus of discoglossoid frogs: its structure and bearing on the systematics of the Anura (Amphibia: Anura). J Zoolog Syst Evol Res 35:179-197.

Haas A. 2003. Phylogeny of frogs as inferred from primarily larval characters (Amphibia: Anura). Cladistics 19:23-89.

Hall BG. 2005. Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. Mol Biol Evol 22:792-802.

Hallam A. 1994. An outline of Phanerozoic biogeography. Oxford University Press, USA.

Hallböök F, Ibáñez CF, and Persson H. 1991. Evolutionary studies of the nerve growth factor family reveal a novel member abundantly expressed in *Xenopus* ovary. Neuron 6:845-858.

Handrigan GR, and Wassersug RJ. 2007. The anuran Bauplan: a review of the adaptive, developmental, and genetic underpinnings of frog and tadpole morphology. Biol Rev 82:1-25.

Hanken J, and Wassersug RJ. 1981. The visible skeleton. A new double-stain technique reveals the native of the "hard" tissues. Funct Photogr 16:22-26.

Hanken J, Jennings DH, and Olsson L. 1997. Mechanistic basis of life-history evolution in anuran amphibians: direct development. Am Zool 37:160-171.

Hasegawa M, Kishino H, and Yano T-a. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160-174.

Hasegawa M, and Kishino H. 1989. Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders. Jpn J Genet 64:243-258.

Hashimoto T, Nakamura Y, Kamishi T, Nakamura F, Adachi J, Okamoto K, and Hasegawa M. 1995. Phylogenetic place of mitochondrion-lacking protozoan *Giardia lamblia*, inferred from amino acid sequences of elongation factor 2. Mol Biol Evol 12:782-793.

Hassanin A, Bonillo C, Bui X, and Cruaud C. 2010. Comparisons between mitochondrial genomes of domestic goat (*Capra hircus*) reveal the presence of numts and multiple sequencing errors. Mitochondrial DNA 21:68-76.

Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their application. Biometrika 57:97-109.

Hay JM, Ruvinsky I, Hedges SB, and Maxson LR. 1995. Phylogenetic relationships of amphibian families inferred from DNA sequences of mitochondrial 12S and 16S ribosomal RNA genes. Mol Biol Evol 12:928-937.

Hayes TB, Collins A, Lee M, Mendoza M, Noriega N, Stuart AA, and Vonk A. 2002. Hermaphroditic, demasculinized frogs after exposure to the herbicide atrazine at low ecologically relevant doses. Proc Natl Acad Sci USA 99:5476-5480.

Hedges BS, Moberg KD, and Maxson LR. 1990. Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationshps. Mol Biol Evol 7:607-633.

Hedges BS. 1992. The number of replicates needed for accurate estimation of the bootstrap p value in phylogenetic studies. Mol Biol Evol 9:366-369.

Hedges SB, and Maxson LR. 1993. A molecular perspective on lissamphibian phylogeny. Herpetol Monogr 7:27-42.

Hedges SB, and Kumar S. 2004. Precision of molecular time estimates. Trends Genet 202:242-247.

Hedges SB, and Kumar S. 2009. Discovering the timetree of life. Pp. 3-18. In Hedges SB, and Kumar S, editors. The Timetree of Life. Oxford University Press, New York.

Hediger M, Romero M, Peng J-B, Rolfs A, Takanaga H, and Bruford E. 2004. The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. Pflug Arch Eur J Phy 447:465-468.

Heled J, and Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. Mol Biol Evol 27:570-580.

Hennig W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana.

Henrici AC. 1994. *Tephrodytes brassicarvalis*, new genus and species (Anura: Pelodytidae), from the Arikareean Cabbage Patch Beds of Montana, USA, and pelodytid-pelobatid relationshps. Ann Carnegie Mus 63:155-183.

Henrici AC. 1998. A new pipoid anuran from the Late Jurasic Morrison formation at Dinosaur National Monument, Utah. J Vert Paleontol 18:321-332.

Higgins D, and Lemey P. 2009. Multiple sequence alignment. Pp. 68-106. In Lemey P, Salemi M, and Vandamme AM, editors. The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, New York.

Hildebrand M. 1968. Anatomical preparations. University of California Press, Berkeley and Los Angeles.

Hillis DM. 1987. Molecular versus morphological approaches to systematics. Annu Rev Ecol Syst 18:23-42.

Hillis DM, Ammerman LK, Dixon MT, and de Sá RO. 1993. Ribosomal DNA and the phylogeny of frogs. Herpetol Monogr 7:1118-1200.

Hillis DM, and Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol 42:182-192.

Hillis DM. 1996. Inferring complex phylogenies. Nature 383:130-131.

Hillis DM. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst Biol 47:3-8.

Hixson JE, Wong TW, and Clayton DA. 1986. Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J Biol Chem 261:2384-2390.

Ho JWK, Adams CE, Lew JB, Matthews TJ, Ng CC, Shahabi-Sirjani A, Tan LH, Zhao Y, Easteal S, Wilson SR, and Jermiin LS. 2006. SeqVis: Visualization of compositional heterogeneity in large alignments of nucleotides. Bioinformatics 22:2162-2163.

Ho SYW, and Jermiin LS. 2004. Tracing the decay of the historical signal in biological sequence data. Syst Biol 53:623-637.

Ho SYW. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds J Avian Biol 38:409-414.

Ho SYW. 2009. An examination of phylogenetic models of substitution rate variation among lineages. Biol Lett 5:421-424.

Ho SYW, and Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst Biol 58:367-380.

Hoegg S, Vences M, Brinkmann H, and Meyer A. 2004. Phylogeny and comparative substitution rates of frogs inferred from sequences of three nuclear genes. Mol Biol Evol 21:1188-1200.

Hof C, Araujo MB, Jetz W, and Rahbek C. 2011. Additive threats from pathogens, climate and land-use change for global amphibian diversity. Nature 480: 516–519.

Holder M, and Lewis PO. 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat Rev Genet 4:275-284.

Hooft RWW, Sander C, Scharf M, and Vriend G. 1996. The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value. Comput Appl Biosci 12:525-529.

Huber KT, Oxelman B, Lott M, and Moulton V. 2006. Reconstructing the evolutionary history of polyploids from multilabeled trees. Mol Biol Evol 23:1784-1791.

Hudson RR. 1983. Properties of a neutral allele model with intragenic recombination. Theor Popul Biol 23:183-201.

Huelsenbeck JP. 1991. When are fossils better than extant taxa in phylogenetic analysis? Syst Biol 40:458-469.

Huelsenbeck JP, and Hillis DM. 1993. Success of phylogenetic methods in the four-taxon case. Syst Biol 42:247-264.

Huelsenbeck JP. 1995a. The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. Mol Biol Evol 12:843-849.

Huelsenbeck JP. 1995b. Performance of phylogenetic methods in simulation. Syst Biol 44:17-48.

Huelsenbeck JP, and Rannala B. 1997. Phylogenetic methods come of age: testing hypotheses in an evoutionary context. Science 276:227-232.

Huelsenbeck JP, and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Huelsenbeck JP, and Rannala B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst Biol 53:904-913.

Huising MO, Stet RJM, Kruiswijk CP, Savelkoul HFJ, and Lidy Verburg-van Kemenade BM. 2003. Molecular evolution of CXC chemokines: extant CXC chemokines originate from the CNS. Trends Immunol 24:306-312.

Hyrien O. 2000. Mechanisms and consequences of replication fork arrest. Biochimie 82:5-17.

Igawa T, Kurabayashi A, Usuki C, Fujii T, and Sumida M. 2008. Complete mitochondrial genomes of three neobatrachians anurans: a case study of divergence time estimation using different data and calibration settings. Gene 407:116-129.

Inger RF. 1967. The development of a phylogeny of frogs. Evolution 21:369-384.

Inoue JG, Miya M, Tsukamoto K, and Nishida M. 2001. Complete mitochondrial DNA sequence of *Conger myriaster* (Teleostei: Anguilliformes): novel gene order for vertebrate mitochondrial genomes and the phylogenetic implications for anguilliform families. J Mol Evol 52:311-320.

Inoue JG, Miya M, Tsukamoto K, and Nishida M. 2004. Mitogenomic evidence for the monophyly of elopomorph fishes (Teleostei) and the evolutionary origin of the leptocephalus larva. Mol Phylogenet Evol 32:274-286.

Ishiguro NB, Miya M, and Nishida M. 2003. Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the "Protacanthopterygii". Mol Phylogenet Evol 27:476-488.

Jacob F. 1977. Evolution and tinkering. Science 196:1161-1166.

Jacobs HT, Elliott DJ, Math VB, and Farquharson A. 1988. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J Mol Biol 202:185-217.

James TY, Litvintseva AP, Vilgalys R, Morgan JAT, Taylor JW, Fisher MC, Berger L, Weldon C, du Preez L, and Longcore JE. 2009. Rapid global expansion of the fungal disease chytridiomycosis into declining and healthy amphibian populations. PLoS Pathogens 5:e1000458.

Janke A, Feldmaier-Fuchs G, Thomas WK, Von-Haeseler A, and Pääbo S. 1994. The marsupial mitochondrial genome and the evolution of placental mammals. Genetics 137:243-256.

Jayaswal V, Robinson J, and Jermiin L. 2007. Estimation of phylogeny and invariant sites under the general Markov model of nucleotide sequence evolution. Syst Biol 56:155-162.

Jayaswal V, Jermiin LS, Poladian L, and Robinson J. 2011. Two stationary nonhomogeneous Markov models of nucleotide sequence evolution. Syst Biol 60:74-86.

Jeffroy O, Brinkmann H, Delsuc F, and Hervé P. 2006. Phylogenomics: the beginning of incongruence? Trends Genet 22:225-231.

Jenkins PA, and Walsh DM. 1993. An Early Jurassic caecilian with limbs. Nature 365:246-250.

Jennings MR, and Hayes MP. 1985. Pre-1900 overharvest of California red-legged frogs (*Rana aurora draytonii*): the inducement for bullfrog (*Rana catesbeiana*) introduction. Herpetologica 41:94-103.

Jermiin LS, Jayaswal V, Ababneh F, and Robinson J. 2008. Phylogenetic model evaluation. Pp. 331-364. In Keith JM, editor. Bioinformatics, Volume I: data, sequence analysis, and evolution. Springer Verlag, Notowa.

Jiang Z, Castoe T, Austin C, Burbrink F, Herron M, McGuire J, Parkinson C, and Pollock D. 2007. Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. BMC Evol Biol 7:123.

Johnson GD, and Patterson C. 1993. Percomorph phylogeny: a survey of acanthomorphs and a new proposal. Bull Mar Sci 52:554–626.

Jukes TH, and Cantor CR. 1969. Evolution of protein molecules. Pp. 21-132. In Munro MN, editor. Mammalian Protein Metabolism. Academic Press, New York.

Jukes TH, and Osawa S. 1993. Evolutionary changes in the genetic code. Comp Biochem Phys B 106:489-494.

Kang D, Miyako K, Kai Y, Irie T, and Takeshige K. 1997. *In vivo* determination of replication origins of human mitochondrial DNA by ligation-mediated polymerase chain reaction. J Biol Chem 272.

Kapitonov V, and Jurka J. 2005. RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons. PLoS Biol 3:e181.

Katoh K, Misawa K, Kuma K-i, and Miyata T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nuc Acids Res 30:3059-3066.

Katoh K, and Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9:286-298.

Kats LB, and Ferrer RP. 2003. Alien predators and amphibian declines: review of two decades of science and the transition to conservation. Divers Distrib 9:99-110.

Kiesecker JM, Blaustein AR, and Belden LK. 2001. Complex causes of amphibian population declines. Nature 410:681-684.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111-120.

Kishino H, Thorne JL, and Bruno WJ. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol Biol Evol 18:352-361.

Kjer KM. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. Mol Phylogenet Evol 4:314-330.

Kluge AG, and Farris JS. 1969. Quantitative phyletics and the evolution of anurans. Syst Zool 18:1-32.

Kluge AG. 1979. Testability and the refutation and corroboration of cladistic hypotheses. Cladistics 13:81-96.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, and Wilson AC. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Ntl Acad Sci USA 86:6196-6200.

Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lieb B, and Halanych KM. 2011. Phylogenomics reveals deep molluscan relationships. Nature 477:452-456.

Köhler J, Vieites DR, Bonett RM, García FH, Glaw F, Steinke D, and Vences M. 2005. New amphibians and global conservation: a boost in species discoveries in a highly endangered vertebrate group. BioScience 55:693-696.

Koonin EV. 2005. Orthologs, paralogs, and evolutionary genomics. Ann Rev Genet 39:309-338.

Kosakovsky Pond SL, Poon AFY, and Frost SD, W. 2009. Estimating selection pressures on alignments of coding sequences. Pp. 419-490. In Lemey P, Salemi M, and Vandamme AM, editors. The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, New York.

Krogh A, Larsson B, von Heijne G, and Sonnhammer ELL. 2011. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567-580.

Kubatko LS, Carstens BC, and Knowles LL. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25:971-973.

Kuhner MK, and Felsenstein J. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol 11:459-469.

Kullback S, and Leibler RA. 1951. On information and sufficiency. Ann Math Stat 22:79-86.

Kumar S, Filipski AJ, Battistuzzi FU, Kosakovsky Pond SL, and Tamura K. 2012. Statistics and truth in phylogenomics. Mol Biol Evol 29:457-472.

Kumazawa Y, Ohta H, and Nishida M. 1995. Variation in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. Mol Biol Evol 12:759-772.

Kunz K. 2003. Krallenfrösche, Zwergkrallenfrösche, Wabenkröten. Pipidae in Natur und Menschenhand. Natur- und Tier-Verlag, Münster.

Kurabayashi A, Usuki C, Mikami N, Fujii T, Yonekawa H, Sumida M, and Hasegawa M. 2006. Complete nucleotide sequence of the mitochondrial genome of a Malagasy poison frog *Mantella madagascariensis*: evolutionary implications on mitochondrial genomes of higher anuran groups. Mol Phylogenet Evol 39:223-236.

Kurabayashi A, Sumida M, Yonekawa H, Glaw F, Vences M, and Hasegawa M. 2008. Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in Mantellid frogs from Madagascar. Mol Biol Evol 25:874-891.

Kurabayashi A, and Sumida M. 2009. PCR Primers for the neobatrachian mitochondrial genome. Curr Herpetol 28:1-11.

Kurabayashi A, Yoshikawa N, Sato N, Hayashi Y, Oumi S, Fujii T, and Sumida M. 2010. Complete mitochondrial DNA sequence of the endangered frog *Odorrana ishikawae* (family Ranidae) and unexpected diversity of mt gene arrangements in ranids. Mol Phylogenet Evol 56:543-553.

Kvist L, Martens J, Nazarenko AA, and Orell M. 2003. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). Mol Biol Evol 20:243-247.

Lahti DC, Johnson NA, Ajie BC, Otto SP, Hendry AP, Blumstein DT, Coss RG, Donohue K, and Foster SA. 2009. Relaxed selection in the wild. Trends Ecol Evol 24:487-496.

Lanfear R, Thomas JA, Welch JJ, Brey T, and Bromham L. 2007. Metabolic rate does not calibrate the molecular clock. Proc Natl Acad Sci USA 104:15388-15393.

Lanfear R, Ho SYW, Love D, and Bromham L. 2010. Mutation rate is linked to diversification in birds. Proc Natl Acad Sci USA 107:20423-20428.

Lanza B, Cei JM, and Crespo E. 1975. Immunological evidence for the specific status of *Discoglossus pictus* Otth, 1837 and *D. sardus* Tschudi, 1837, with notes on the families Discoglossidae Günther, 1858 and Bombinidae Fitzinger, 1826 (Amphibia: Salientia). Monit Zool Ital (N. S.) 9:153-162.

Lartillot N, and Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 21:1095-1109.

Lathrop A. 1997. Taxonomic review of the megophryid frogs (Anura: Pelobatoidea). Asiatic Herpetol Res 7:68-79.

Laurent RF. 1979. Esquisse d'une phylogenèse des anoures. Bull Soc Zool France 104:397-422.

Laurent RF. 1985. Sur la classification et la nomenclature des Amphibiens. Alytes 4:119-120.

Laurent RF. 1986. Sous classe des lissamphibiens (Lissamphibia). Pp. 594-797. In Grassé P, and Delsol M, editors. Traité de zoologie. Anatomie, systematique, biologie. Masson, Paris.

Laurin M, and Reisz R. 1997. A new perspective on tetrapod phylogeny. In Sumida SS, and Martin KL, editors. Amniote origins: completing the transition to land. Academic Press, New York.

Laurin M. 1998. The importance of global parsimony and historical bias in understanding tetrapod evolution. Part I. Systematics, middle ear evolution and jaw suspension. Ann Sci Nat Zool 19:1-42.

Lee WJ, and Kocher TD. 1995. Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. Genetics 139:873-887.

Lemey P, and Posada D. 2009. Molecular clock analysis. Pp. 362-377. In Lemey P, Salemi M, and Vandamme A-M, editors. The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, New York.

Lemmon AR, and Moriarty EC. 2004. The importance of proper model assumption in Bayesian phylogenetics. Syst Biol 53:265-277.

Levinger L, Giegé R, and Florentz C. 2003. Pathology-related substitutions in human mitochondrial $tRNA^{Ile}$ reduce precursor 3' end processing efficiency *in vitro*. Nuc Acids Res 31:1904-1912.

Levinson G, and Gutman GA. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203-221.

Li JT, Che J, Murphy RW, Zhao H, Zhao EM, Rao DQ, and Zhang YP. 2009. New insights to the molecular phylogenetics and generic assessment in the Rhacophoridae (Amphibia: Anura) based on five nuclear and three mitochondrial genes, with comments on the evolution of reproduction. Mol Phylogenet Evol 53:509-522.

Li P, and Bousquet J. 1992. Relative rate test for nucleotide substitutions between two lineages. Mol Biol Evol 9:1185-1189.

Li S, Pearl D, and Doss H. 2000. Phylogenetic tree reconstruction using Markov chain Monte Carlo. J Am Stat Assoc 95:493-508.

Li WH, Wu CI, and Luo CC. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150-174.

Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K, and Engel A. 2003. Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. J Biol Chem 278:21655-21662.

Liu L, and Pearl DK. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst Biol 56:504-514.

Liu L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24:2542-2543.

Liu ZQ, Wang YQ, and Su B. 2005. The mitochondrial genome organization of the rice frog, *Fejervarya limnocharis* (Amphibia: Anura): a new gene order in the vertebrate mtDNA. Gene 346:145-151.

Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, and Larkum AWD. 1992. Substitutional bias confounds inference of cyanelle origins from sequence data. J Mol Evol 34:153-162.

Loomis WF, and Smith DW. 1990. Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. Proc Natl Acad Sci USA 87:9093-9097.

Lupi R, de Meo PDO, Picardi E, D'Antonio M, Paoletti D, Castrignanò T, Pesole G, and Gissi C. 2010. MitoZoa: A curated mitochondrial genome database of metazoans for comparative genomics studies. Mitochondrion 10:192-199.

Lynch JD. 1973. The transition from archaic to advanced frogs. Pp. 133-182. In Vial JL, editor. Evolutionary biology of the anurans: contemporary research on major problems. University of Missouri Press, Columbia.

Lynch M. 1971. Evolutionary relationships, osteology, and zoogeography of leptodactylid frogs. Misc Pub Mus Nat Hist Univ Kansas 53:1-238.

Lynch M. 2007. The origins of genome architecture. Sinauer Associates Inc., Sunderland, Massacusetts.

Lytton J. 2007. Na^{+}/Ca^{2+} exchangers: three mammalian gene families control Ca^{2+} transport Biochem J 406:365-382.

Mabuchi K, Miya M, Satoh TP, Westneat MW, and Nishida M. 2004. Gene rearrangements and evolution of tRNA pseudogenes in the mitochondrial genome of the Parrotfish (Teleostei: Perciformes: Scaridae). J Mol Evol 59:287-297.

Macey JR, Larson A, Ananjeva NB, Fang Z, and Papenfuss TJ. 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol Biol Evol 14:91-104.

Macey JR, Schulte JA, 2nd, Larson A, and Papenfuss TJ. 1998. Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol Biol Evol 15:71-75.

Macey JR, Papenfuss TJ, Kuehl JV, Fourcade HM, and Boore JL. 2004. Phylogenetic relationships among amphisbaenian reptiles based on complete mitochondrial genomic sequences. Mol Phylogenet Evol 33:22-31.

MacKay SLD, Olivo PD, Laipis PJ, and Hauswirth WW. 1986. Template-directed arrest of mammalian mitochondrial DNA synthesis. Mol Cell Biol 6:1261-1267.

Maddison WP, and Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst Biol 55:21-30.

Magallón S. 2010. Using fossils to break long branches in molecular dating: a comparison of relaxed clocks applied to the origin of angiosperms. Syst Biol 59:384-399.

Maglia AM, Púgener LA, and Trueb L. 2001. Comparative development of anurans: using phylogeny to understand ontogeny. Am Zool 41:538-551.

Marjanović D, and Laurin M. 2007. Fossils, molecules, divergence times, and the origin of lissamphibians. Syst Biol 56:369-388.

Marjanović D, and Laurin M. 2008. A reevaluation of the evidence supporting an unorthodox hypothesis on the origin of extant amphibians. Contrib Zool 77:149–199.

Marjanović D, and Laurin M. 2009. The origin(s) of modern amphibians: a commentary. Evol Biol 36:336-338.

Márquez R, Pargana JM, and Crespo EG. 2001. Acoustic competition in male *Pelodytes ibericus* (Anura: Pelodytidae): interactive playback tests. Copeia 2001:1142-1150.

Marsh DM, and Trenham PC. 2001. Metapopulation dynamics and amphibian conservation. Conserv Biol 15:40-49.

Martens PA, and Clayton DA. 1979. Mechanism of mitochondrial DNA replication in mouse L-cells: localization and sequene of the light-strand origin of replication. J Mol Biol 135:327-351.

Martin AP, and Burg TM. 2002. Perils of paralogy: using HSP70 genes for inferring organismal phylogenies. Syst Biol 51:570-587.

Martín C, and Sanchíz FB. 2010. Lisanfos KMS. Version 1.2. Online reference accessible at http://www.lisanfos.mncn.csic.es/. Museo Nacional de Ciencias Naturales, CSIC. Madrid, Spain.

Mau B, and Newton MA. 1997. Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. J Comput Graph Stat 6:122-131.

Maxson LR, and Szymura JM. 1984. Relationships among discoglossid frogs: an albumin perspective. Amphibia-Reptilia 5:245-252.

Maxson R, Cohn R, Kedes L, and Mohun T. 1983. Expression and organization of histone genes. Annu Rev Genet 17:239-277.

Meiklejohn CD, Montooth KL, and Rand DM. 2007. Positive and negative selection on the mitochondrial genome. Trends Genet 23:259-263.

Menzies JI. 1967. An ecological note on the frog *Pseudhymenochirus merlini* Chabanaud in Sierra Leone. J W Afr Sci Ass 12:23-28.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, and Teller E. 1953. Equations of state calculations by fast computing machines. J Chem Phys 21:1087-1092.

Meyer A. 1993. Evolution of mitochondrial DNA in fishes. Pp. 1-38. In Hochachaka PW, Mommsen TP, editors. Biochemistry and molecular biology of fishes vol. 2. Elsevier Science Publishers, Amsterdam.

Milner AR. 1988. The relationships and origin of living amphibians. Pp. 59-102. In Benton MJ, editor. The phylogeny and classification of the tetrapods. Clarendon Press, Oxford.

Milner AR. 1993. Amphibian-grade Tetrapoda. Pp. 665-679. In Benton MJ, editor. The Fossil Record 2. Chapman & Hall, London.

Mindell DP, Sorenson MD, and Dimcheff DE. 1998. Multiple independent origins of mitochondrial gene order in birds. Proc Natl Acad Sci USA 95:10693-10697.

Miya M, and Nishida M. 2000. Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. Mol Phylogenet Evol 17:437-455.

Miya M, Kawaguchi A, and Nishida M. 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol Biol Evol 18:1993-2009.

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, and Akira K. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol Phylogenet Evol 26:121-138.

Moepps B, Braun M, Knöpfle K, Dillinger K, Knöchel W, and Gierschik P. 2000. Characterization of a *Xenopus laevis* CXC chemokine receptor 4: implications for hemato-poietic cell development in the vertebrate embryo. Eur J of Immunol 30:2924-2934.

Montgelard C, Catzeflis FM, and Douzery E. 1997. Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome b and 12S rRNA mitochondrial sequences. Mol Biol Evol 14:550-559.

Montoya J, Christianson T, Levens D, Rabinowitz M, and Attardi G. 1982. Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. Proc Natl Acad Sci USA 79:7195-7199.

Montoya J, Gaines GL, and Attardi G. 1983. The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell 34:151-159.

Moore WS. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49: 718-726.

Moritz C, and Brown WM. 1986. Tandem duplications of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science 233:1425-1427.

Moritz C, and Brown WM. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc Natl Acad Sci USA 84:7183-7187.

Moritz C, Dowling TE, and Brown WM. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu Rev Ecol Syst 18:269-292.

Mueller RL, Macey JR, Jaekel M, Wake DB, and Boore JL. 2004. Morphological homoplasy, life history evolution, and historical biogeography of plethodontid salamanders inferred from complete mitochondrial genomes. Proc Natl Acad Sci USA 101:13820-13825.

Mueller RL, and Boore JL. 2005. Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. Mol Biol Evol 22:2104-2112.

Mueller RL. 2006. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. Syst Biol 55:289-300.

Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res-Fund Mol M 1:2-9.

Müller R, and Scheer U. 1970. Sound spectroscopy of calls in the clawed toad, *Xenopus laevis*. Experientia 26:435-436.

Nabholz B, Mauffrey J-F, Bazin E, Galtier N, and Glemin S. 2008. Determination of mitochondrial genetic diversity in mammals. Genetics 178:351-361.

Nei M, and Kumar S. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

Nelson JS. 2006. Fishes of the world. John Wiley & Sons, Inc., Hoboken, New Jersey.

Nesnidal MP, Helmkampf M, Bruchhaus I, and Hausdorf B. 2010. Compositional heterogeneity and phylogenomic inference of metazoan relationships. Mol Biol Evol 27:2095-2104.

Nielsen R, and Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929-936.

Noble GK. 1922. The phylogeny of the Salientia. I. The osteology and the thigh musculature: Their bearing on classification and phylogeny. Bull Amer Mus Nat Hist 46:1-87.

Noble GK. 1931. The biology of Amphibia. Dover Publishing, New York.

Nylander JAA, Wilgenbusch JC, Warren DL, and Swofford DL. 2008. AWTY (are we there yet?): A system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581-583.

Ogg JG, Ogg G, and Gradstein FM. 2008. The concise geologic time scale. Cambridge University Press, Cambridge.

Ojala D, Merkel C, Gelfand R, and Attardi G. 1980. The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. Cell 22:393-403.

Ojala D, Montoya J, and Attardi G. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470-474.

Olsson L, and Hanken J. 1996. Cranial neural-crest migration and chondrogenic fate in the oriental fire-bellied toad *Bombina orientalis*: defining the ancestral pattern of head development in anuran amphibians. J Morphol 229:105-120.

Orton GL. 1953. The systematics of vertebrate larvae. Syst Zool 2:63-75.

Orton GL. 1957. The bearing of larval evolution on some problems in frog classification. Syst Zool 6:79-86.

Osawa S, Jukes TH, Watanabe K, and Muto A. 1992. Recent evidence for evolution of the genetic code. Microbiol Rev 56:229–264.

Otto SP. 2007. The evolutionary consequences of polyploidy. Cell 131:452-462.

Pääbo S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci USA 86:1939-1943.

Pabijan M, Christina Spolsky C, Thomas Uzzell T, and Szymura JM. 2008. Comparative analysis of mitochondrial genomes in *Bombina* (Anura: Bombinatoridae). J Mol Evol 67:246-256.

Pagani I, Liolios K, Jansson J, Chen I-MA, Smirnova T, Nosrat B, Markowitz VM, and Kyrpides NC. 2012. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nuc Acids Res 40:D571-D579.

Page RDM, and Holmes EC. 1998. Molecular evolution: a phylogenetic approach. Blackwell Publishing Ltd., Oxford.

Pagel M. 1997. Inferring evolutionary processes from phylogenies. Zool Scr 26:331-348.

Pagel M, Meade A, and Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. Syst Biol 53:673-684.

Pagel M, Venditti C, and Meade A. 2006. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science 314:119-121.

Pagel M, and Meade A. 2008. Modelling heterotachy in phylogenetic inference by reversible-jump Markov chain Monte Carlo. Philos Trans R Soc B 363:3955-3964.

Palumbi S, Martin A, Romano S, McMillan WO, Stice L, and Grabowski G. 1991. The simple fool's guide to PCR. Department of Zoology, Kewalo Marine Laboratory, Univerisy of Hawaii.

Palumbi SR. 1989. Rates of molecular evolution and the function of nucleotide positions free to vary. J Mol Evol 29:180-187.

Parker WK. 1871. On the structure and development of the skull of the common frog (*Rana temporaria*, L.). Philos Trans R Soc Lond 161:137-211.

Pattengale N, Alipour M, Bininda-Emonds O, Moret B, and Stamatakis A. 2009. How many bootstrap replicates are necessary? Pp. 184-200. In Batzoglou S, editor. Research in Computational Molecular Biology. Springer, Berlin and Heidelberg.

Pearl CA, Cervantes M, Chan M, Ho U, Shoji R, and Thomas EO. 2000. Evidence for a mate-attracting chemosignal in the dwarf African clawed frog *Hymenochirus*. Horm Behav 38:67-74.

Perna NT, and Kocher TD. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353-358.

Philippe H. 2000. Long branch attraction and protist phylogeny. Protist 51:307-316.

Philippe H, and Germot A. 2000. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. Mol Biol Evol 17:830-834.

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, and Casane D. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol Biol Evol 21:1740-1752.

Philippe H, Zhou Y, Brinkmann H, Rodrigue N, and Delsuc F. 2005. Heterotachy and long-branch attraction in phylogenetics. BMC Evol Biol 5:50.

Phillips MJ, Delsuc F, and Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. Mol Biol Evol 21:1455-1458.

Poe S, and Swofford DL. 1999. Taxon sampling revisited. Nature 389:299-300.

Poe S. 2003. Evaluation of the strategy of long-branch subdivision to improve the accuracy of phylogenetic methods. Syst Biol 52:423-428.

Posada D, and Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

Posada D, and Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol 53:793-808.

Posada D. 2008. jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253-1256.

Posada D. 2009. Selecting models of evolution. Pp. 345-361. In Lemey P, Salemi M, and Vandamme AM, editors. The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, New York.

Pruitt KD, Tatusova T, Klimke W, and Maglott DR. 2009. NCBI Reference Sequences: current status, policy and new initiatives. Nuc Acids Res 37(Database issue): D32-D36.

Púgener LA, Maglia AM, and Trueb L. 2003. Revisiting the contribution of larval characters to an analysis of phylogenetic relationships of basal anurans. Zool J Linn Soc 139:129-155.

Puritz JB, Addison JA, and Toonen RJ. 2012. Next-generation phylogeography: a targeted approach for multilocus sequencing of non-model organisms. PLoS ONE 7:e34241.

Pybus OG, and Shapiro B. 2009. Natural selection and adaptation of molecular sequences. Pp. 407-418. In Lemey P, Salemi M, and Vandamme AM, editors. The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, New York.

Pyron RA, and Wiens JJ. 2011. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol Phylogenet Evol 61:543-583.

Pyron RA. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Syst Biol 60:466-481.

Quednau B, Nicoll D, and Philipson K. 2004. The sodium/calcium exchanger family—SLC8. Pflug Arch Eur J Phy 447:543-548.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Rabb GB. 1960. On the unique sound production of the Surinam toad, *Pipa pipa*. Copeia 4:368-369.

Rabb GB, and Rabb MS. 1963. On the behavior and breeding biology of the African pipid frog *Hymenochirus boettgeri*. Z Tierpsychol 20:215-241.

Rabb GB. 1969. Fighting frogs. Brookfield Bandarlog 37:4-5.

Rabb GB. 1973. Evolutionary aspects of the reproductive behavior of frogs. Pp. 213-227. In Vial JL, editor. Evolutionary biology of the anurans: contemporary research on major problems. University of Misouri Press, Columbia.

Rage J, and Roček Z. 1989. Redescription of *Triadobatrachus massinoti* (Piveteau, 1936) an anuran amphibian from the early Triassic. Palaeontogr Abt A 206:1-16.

Rage JC, and Janvier P. 1982. Le problème de la monophylie des amphibiens actuels, à la lumière des nouvelles donnees sur les affinités des tétrapodes. Geobios 6:65-83.

Rambaut A, and Drummond AJ. 2009. Tracer v. 1.5, Available from http://tree.bio.ed.ac.uk/software/tracer/.

Rand D. 2001. Mitochondrial genomics flies high. Trends Ecol Evol 16:2-4.

Rannala B, and Yang Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J Mol Evol 43:304-311.

Rannala B, and Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645–1656.

Rannala B, Zhu T, and Yang Z. 2012. Tail paradox, partial identifiability and influential priors in Bayesian branch length inference. Mol Biol Evol 29:325-335.

Rawson PD, and Hilbish TJ. 1995. Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. Mol Biol Evol 15:893-901.

Recuero E, Canestrelli D, Vörös J, Szabó K, Poyarkov NA, Arntzen JW, Crnobrnja-Isailovic J, Kidov AA, Cogălniceanu D, Caputo FP, Nascetti G, and Martínez-Solano I. 2011. Multilocus species tree analyses resolve the radiation of the widespread *Bufo bufo* species group (Anura, Bufonidae). Mol Phylogenet Evol 62:71-86.

Reeves JH. 1992. Heterogeneity in the subsitution process of amino acid sites of proteins coded by mitochondrial DNA. J Mol Evol 35:17-31.

Reig OA. 1958. Proposiciones para una nueva macrosistemática de los anuros (nota preliminar). Physis 21:109-118.

Ren Z, Zhu B, Ma E, Wen J, Tu T, Cao Y, Hasegawa M, and Zhong Y. 2009. Complete nucleotide sequence and gene arrangement of the mitochondrial genome of the crab-eating frog *Fejervarya cancrivora* and evolutionary implications. Gene 441:148-155.

Reyes A, Gissi C, Pesole G, and Saccone C. 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol Biol Evol 15:957-966.

Richter C, Park JW, and Ames BN. 1988. Normal oxidative damage to mitochondrial and nuclear DNA is extensive. Proc Natl Acad Sci USA 85:6465-6467.

Ridewood WG. 1897. On the structure and development of the hyobranchial skeleton and larynx in *Xenopus* and *Pipa*, with remarks on the affinities of the Aglossa. J Linn Soc Lond 26:53-128.

Ridewood WG. 1900. On the hyobranchial skeleton and larynx of the new aglossal toad, *Hymenochirus boettgeri*. J Linn Soc Lond 27:454-460.

Ritland RM. 1955. Studies on the post-cranial morphology of *Ascaphus truei*. J Morphol 97:215–282.

Robinson M, Gouy M, Gautier C, and Mouchiroud D. 1998. Sensitivity of the relative-rate test to taxonomic sampling. Mol Biol Evol 15:1091-1098.

Robinson-Rechavi M, and Huchon D. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics 16:296-297.

Roček Z, and Vesely M. 1989. Development of the ethmoidal structures of the endocranum in the anuran *Pipa pipa*. J Morphol 200:300–319.

Roček Z. 2000. Mesozoic anurans. Pp. 1295-1331. In Heatwole H, and Carroll RL, editors. Amphibian biology. Surrey Beatty, Chipping Norton, Australia.

Roček Z, and Rage JC. 2000. Tertiary anura of Europe, Asia, Africa, Asia, North America and Australia. Pp. 1332-1387. In Heatwole H, and Carroll RL, editors. Amphibian Biology. Surrey Beatty, Chipping Norton, Australia.

Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang FB, and Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst Biol 56:389-399.

Roe BA, Ma DP, Wilson RK, and Wong JF. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J Biol Chem 260:9759-9774.

Roelants K, and Bossuyt F. 2005. Archaeobatrachian paraphyly and Pangaean diversification of crown-group frogs. Syst Biol 54:111-126.

Roelants K, Gower DJ, Wilkinson M, Loader S, Biju SD, Guillaume K, Moriau L, and Bossuyt F. 2007. Global patterns of diversification in the history of modern amphibians. Proc Natl Acad Sci USA 104:887-892.

Roelants K, Haas A, and Bossuyt F. 2011. Anuran radiations and the evolution of tadpole morphospace. Proc Natl Acad Sci USA 108:8731-8736.

Rogers JS. 1997. On the consistency of maximum likelihood estimation of phylogenetic trees from nucleotide sequences. Systems Biol 46:354-357.

Rokas A, and Holland PWH. 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol 15:454-459.

Rokas A, and Carroll SB. 2005. More genes or more taxa? the relative contribution of gene number and taxon number to phylogenetic accuracy. Mol Biol Evol 22:1337-1344.

Romiguier J, Figuet E, Galtier N, Douzery EJP, Boussau B, Dutheil JY, and Ranwez V. 2012. Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. PLoS ONE 7:e33852.

Ronquist F, and Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Ronquist F, van der Mark P, and Huelsenbeck JP. 2009. Bayesian phylogenetic analysis using MrBayes. Pp. 210-266. In Lemey P, Salemi M, and Vandamme AM, editors. The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, New York.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, and Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539-542.

Rosenberg MS, and Kumar S. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proc Natl Acad Sci USA 98:10751-10756.

Rossmanith W, Tullo A, Potushak T, Karwan R, and Sbis E. 1995. Human mitochondrial tRNA processing. J Biol Chem 270:12885-12891.

Rothstein R, Michel B, and Gangloff S. 2000. Replication fork pausing and recombination or "gimme a break". Genes Dev 14:1-10.

Ruffalo M, LaFramboise T, and Koyutürk M. 2011. Comparative analysis of algorithms for next-generation sequencing read alignment. Bioinformatics 27:2790-2796.

Russo CAM, Takezaki N, and Nei M. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol Biol Evol 13:525-536.

Ruta M, Coates MI, and Quicke DLJ. 2003. Early tetrapod relationships revisited. Biol Rev 78:251-345.

Ruta M, and Coates MI. 2007. Dates, nodes and character conflict: addressing the lissamphibian origin problem. J Syst Palaeontol 5:69-122.

Ruvinsky I, and Maxson LR. 1996. Phylogenetic relationships among bufonoid frogs (Anura: Neobatrachia) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol 5:533-547.

Rzhetsky A, and Nei M. 1992. A simple method for estimating and testing minimum-evolution trees. Mol Biol Evol 9:945-967.

Saitou N, and Nei M. 1987. The Neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

Sambrook J, Fritsch EF, and Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

San Mauro D, García-París M, and Zardoya R. 2004a. Phylogenetic relationships of discoglossid frogs (Amphibia: Anura: Discoglossidae) based on complete mitochondrial genomes and nuclear genes. Gene 343:357-366.

San Mauro D, Gower DJ, Oommen OV, Wilkinson M, and Zardoya R. 2004b. Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1. Mol Phylogenet Evol 33:413-427.

San Mauro D, Vences M, Alcobendas M, Zardoya R, and Meyer A. 2005. Initial diversification of living amphibians predated the breakup of Pangaea. Am Nat 165:590-599.

San Mauro D, Gower DJ, Zardoya R, and Wilkinson M. 2006. A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. Mol Biol Evol 23:227-234.

San Mauro D, Gower DJ, Massingham T, Wilkinson M, Zardoya R, and Cotton JA. 2009. Experimental design in caecilian systematics: Phylogenetic information of mitochondrial genomes and nuclear *rag1*. Syst Biol 58:425-438.

San Mauro D. 2010. A multilocus timescale for the origin of extant amphibians. Mol Phylogenet Evol 56:554-561.

San Mauro D, and Agorreta A. 2010. Molecular systematics: a synthesis of the common methods and the state of knowledge. Cell Mol Biol Lett 15:311-341.

San Mauro D, Gower DJ, Cotton JA, Zardoya R, Wilkinson M, and Massingham T. 2012. Experimental design in phylogenetics: testing predictions from expected information. Syst Biol in press.

Sanchíz FB. 1998. Encyclopedia of paleoherpetology. Pfeil, München.

Sanderson MJ, and Donoghue MJ. 1989. Patterns of variation in levels of homoplasy. Evolution 43:1781-1795.

Sanderson MJ. 1997. Nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol 14:1218-1231.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: apenalized likelihood approach. Mol Biol Evol 19:101-109.

Sano N, Kurabayashi A, Fujii T, Yonekawa H, and Sumida M. 2004. Complete nucleotide sequence and gene rearrangement of the mitochondrial genome of the bell-ring frog, *Buergeria buergeri* (family Rhacophoridae). Genes Genet Syst 79:151-163.

Sano N, Kurabayashi A, Fujii T, Yonekawa H, and Sumida M. 2005. Complete nucleotide sequence of the mitochondrial genome of Schlegel's tree frog *Rhacophorus schlegelii* (family Rhacophoridae): duplicated control regions and gene rearrangements. Genes Genet Syst80:213-224.

Sarich VM, and Wilson AC. 1973. Generation time and genomic evolution in primates. Science 179:1144-1147.

Satoh TP, Miya M, Endo H, and Nishida M. 2006. Round and pointed-head grenadier fishes (Actinopterygii: Gadiformes) represent a single sister group: evidence from the complete mitochondrial genome sequences. Mol Phylogenet Evol 40:129-138.

Savage JM. 1973. The geographic distribution of frogs: patterns and predictions. Pp. 351-445. In Vial JL, editor. Evolutionary biology of the anurans: contemporary research on major problems. Columbia University Press, Columbia.

Savage RM. 1965. External stimulus for the natural spawning of *Xenopus laevis*. Nature 205:618-619.

Scarpulla RC. 2008. Transcriptional paradigms in mammalian mitochondrial biogenesis and function. Physiol Rev 88:611-638.

Schatz DG, Oettinger MA, and Baltimore D. 1989. The V(D)J recombination activating gene, RAG-1. Cell 59:1035-1048.

Schierup MH, and Hein J. 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics 156:879-891.

Schmidt H. 2009. Testing tree topologies. Pp. 381-403. In Lemey P, Salemi M, and Vandamme AM, editors. The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, New York.

Schmidt H, and von Haeseler A. 2009. Phylogenetic inference using maximum likelihood methods. Pp. 181-207. In Lemey P, Salemi M, and Vandamme AM, editors. The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, New York.

Schneider A, and Ebert D. 2004. Covariation of mitochondrial genome size with gene lengths: evidence for gene length reduction during mitochondrial evolution. J Mol Evol 59:90-96.

Schoch RR, and Carroll RL. 2003. Ontogenetic evidence for the Paleozoic ancestry of salamanders. Evol Dev 5:314-324.

Schoch RR, and Milner AR. 2004. Structure and implications of theories on the origin of lissamphibians. Pp. 345-377. In Arratia G, Wilson MVH, and Cloutier R, editors. Recent advances in the origin and early radiation of vertebrates, Pfeil, München.

Schultz DW, and Yarus M. 1994. tRNA structure and ribosomal function. II. Interaction between anticodon helix and other tRNA mutations. J Mol Biol 235:1395-1405.

Schwarz G. 1978. Estimating the dimension of a model. Ann Stat 6:461-464.

Shadel GS, and Clayton DA. 1997. Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem 66:409-435.

Shao R, Dowton M, Murrell A, and Barker SC. 2003. Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. Mol Biol Evol 20:1612-1619.

Sheffield NC, Hiatt KD, Valentine MC, Song H, and Whiting MF. 2010. Mitochondrial genomics of Orthoptera using MOSAS. Mitochondrial DNA 21:87-104.

Shields GF, and Kocher TD. 1991. Phylogenetic relationships of north american ursids based on analysis of mitochondrial DNA. Evolution 45:218-221.

Shimodaira H, and Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114-1116.

Shimodaira H, and Hasegawa M. 2001. CONSEL: For assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246-1247.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol 51:592-508.

Shubin NH, and Jenkins FA. 1995. An early Jurassic jumping frog. Nature 377:4952.

Sigurdsen T, and Green DM. 2011. The origin of modern amphibians: a re-evaluation. Zool J Linn Soc 162: 457-469.

Simmons MP, Ochoterena H, and Freudenstein JV. 2002. Amino acid vs. nucleotide characters: challenging preconceived notions. Mol Phylogenet Evol 24:78-90.

Simmons MP, Carr TG, and O'Neill K. 2004. Relative character-state space, amount of potential phylogenetic information, and heterogeneity of nucleotide and amino acid characters. Mol Phylogenet Evol 32:913-926.

Simmons MP. 2012. Radical instability and spurious branch support by likelihood when applied to matrices with non-random distributions of missing data. Mol Phylogenet Evol 62:472-484.

Simpson GG. 1944. Tempo and mode in evolution. Columbia University Press, New York.

Singer GAC, and Hickey DA. 2000. Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. Mol Biol Evol 17:1581-1588.

Smith SA, and Donoghue MJ. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322:86-89.

Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SCS, Rouse GW, Giribet G, and Dunn CW. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature 480: 364–367.

Sober E. 1988. Reconstructing the past: parsimony, evolution, and inference. MIT Press, Cambridge, Massachusetts.

Sokal RR, and Sneath PHA. 1963. Numerical taxonomy. WH Freeman, San Francisco.

Sokol O. 1975. The phylogeny of anuran larvae: a new look. Copeia 1975:1-23.

Sokol OM. 1977. The free swimming *Pipa* larvae, with a review of pipid larvae and pipid phylogeny (Anura: Pipidae). J Morphol 154:357-425.

Soltis PS, Soltis DE, Savolainen V, Crane PR, and Barraclough TG. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. Proc Natl Acad Sci USA 99:4430-4435.

Špinar ZV. 1972. Tertiary frogs from central Europe, Springer, The Hague.

Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, de Jong WW, and Stanhope MJ. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Mol Biol Evol 18:132-143.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Stamatakis A, Blagojevic F, Nikolopoulos D, and Antonopoulos C. 2007. Exploring new search algorithms and hardware for phylogenetics: RAxML meets the IBM cell. J VLSI Signal Proc 48:271-286.

Starrett PH. 1973. Evolutionary patterns in larval morphology. Pp. 251-271. In Vial JL, editor. Evolutionary biology of the anurans: contemporary research on major problems. University of Misouri Press, Columbia.

Stephenson NG. 1951. Observations on the development of the amphicoelous frogs, *Leiopelma* and *Ascaphus*. J Linn Soc Lond 42:18-28.

Stiassny MLJ. 1986. The limits and relationships of the acanthomorph teleosts. J Zool 1:411-460.

Strimmer K, and von Haeseler A. 1996. Quartet-puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964-969.

Stuart SN, Chanson JS, Cox NA, Young BE, Rodrigues ASL, Fischman DL, and Waller RW. 2004. Status and trends of amphibian declines and extinctions worldwide. Science 306:1783-1786.

Stuart SN, Hoffmann M, Chanson JS, Cox NA, Berridge RJ, Ramani P, and Young BE. 2008. Threatened amphibians of the world. Lynx Edicions, Barcelona.

Su X, Wu X-B, Yan P, Cao S-Y, and Hu Y-L. 2007. Rearrangement of a mitochondrial tRNA gene of the concave-eared torrent frog, *Amolops tormotus*. Gene 394:25-34.

Sueur J, Aubin T, and Simonis C. 2008. Seewave: A free modular tool for sound analysis and synthesis. Bioacoustics 18:213–226.

Sullivan J, and Joyce P. 2005. Model selection in phylogenetics. Annu Rev Ecol Evol Syst 36:445-466.

Sumida M, Kanamori Y, Kaneda H, Kato Y, Nishioka M, Hasegawa M, and Yonekawa H. 2001. Complete nucleotide sequence and gene rearrangement of the mitochondrial genome of the Japanese pond frog *Rana nigromaculata*. Genes Genet Syst 76:311-325.

Sun C, Shepard DB, Chong RA, López Arriaza J, Hall K, Castoe TA, Feschotte C, Pollock DD, and Mueller RL. 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol Evol 4:168-183.

Susko E. 2010. First-order correct bootstrap support adjustments for splits that allow hypothesis testing when using maximum likelihood estimation. Mol Biol Evol 27:1621-1629.

Suzuki Y, Glazko GV, and Nei M. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc Natl Acad Sci USA 99:16138-16143.

Swofford DL, and Olsen GJ. 1990. Phylogenetic reconstruction. Pp. 411-501. In Hillis DM, and Moritz C, editors. Molecular systematics. Sinauer Associates, Suderland, Massachusetts.

Swofford DL, Olsen GJ, Waddell PJ, and Hillis DM. 1996. Phylogenetic inference. Pp. 407-514. In Hillis DM, Moritz C, and Mable BK, editors. Molecular systematics. Sinauer Associates, Suderland, Massachusetts.

Swofford DL. 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods).

Talavera G, and Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignmets. Syst Biol 56:564-577.

Tamura K. 1994. Model selection in the estimation of the number of nucleotide substitutions. Mol Biol Evol 11:154-157.

Tamura K, and Kumar S. 2002. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol Biol Evol 19:1727-1736.

Tateno Y, Takezaki N, and Nei M. 1994. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when subtitution rate varies with site. Mol Biol Evol 11:261-277.

Telford MJ, and Copley RR. 2011. Improving animal phylogenies with genomic data. Trends Genet 27:186-195.

Tettamanti G, Cattaneo AG, Gornati R, de Eguileor M, Bernardini G, and Binelli G. 2010. Phylogenesis of brain-derived neurotrophic factor (BDNF) in vertebrates. Gene 450:85-93.

Thatcher TH, and Gorovsky MA. 1994. Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nuc Acids Res 22:174-179.

Thierry-Mieg D, and Thierry-Mieg J. 2006. AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol 7:S12.

Thorne JL, Kishino H, and Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol 15:1647-1657.

Thorne JL, and Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst Biol 51:689-702.

Thyagarajan B, Padua RA, and Campbell C. 1996. Mammalian mitochondria possess homologous DNA recombination activity. J Biol Chem 271:27536-27543.

Townsend DS, Stewart MM, Pough FH, and Brussard PF. 1981. Internal fertilization in an oviparous frog. Science 212:469-471.

Townsend JP. 2007. Profiling phylogenetic informativeness. Syst Biol 56:222-231.

Townsend JP, López-Giráldez F, and Friedman R. 2008. The phylogenetic informativeness of nucleotide and amino acid sequences for reconstructing the vertebrate tree. J Mol Evol 67:437-447.

Townsend JP, and López-Giráldez F. 2010. Optimal selection of gene and ingroup taxon sampling for resolving phylogenetic relationships. Syst Biol 59:446-457.

Townsend TM, Mulcahy DG, Noonan BP, Sites JWJ, Kuczynski CA, Wiens JJ, and Reeder TW. 2011. Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation. Mol Phylogenet Evol 61:363-380.

Trueb L, and Cloutier R. 1991. Toward an understanding of the amphibians: two centuries of systematic history. Pp. 175-193. In Schultze HP, and Trueb L, editors. Origins of the major groups of tetrapods: controversies and consensus. Cornell University Press, Ithaca.

Trueb L, Ross CF, and Smith R. 2005. A new pipoid anuran from the late Cretaceous of South Africa. J Vertebr Paleontol 25:533-547.

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, and Zouros E. 2005. Widespread recombination in published animal mtDNA sequences. Mol Biol Evol 22:925-933.

Tuffley C, and Steel M. 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bull Math Biol 59:581-607.

Turner PC, Bagenal EB, Vlad MT, and Woodland HR. 1988. The organisation and expresion of histone genes from *Xenopus borealis*. Nuc Acids Res 16:3471-3485.

Vallin G, and Laurin M. 2004. Cranial morphology and affinities of *Microbrachis*, and a reappraisal of the phylogeny and lifestyle of the first amphibians. J Vertebr Paleontol 24:56-72.

van der Meijden A, Vences M, Hoegg S, and Meyer A. 2005. A previously unrecognized radiation of ranid frogs in Southern Africa revealed by nuclear and mitochondrial DNA sequences. Mol Phylogenet Evol 37:674-685.

van der Meijden A, Boistel R, Gerlach J, Ohler A, Vences M, and Meyer A. 2007a. Molecular phylogenetic evidence for paraphyly of the genus *Sooglossus*, with the description of a new genus of Seychellean frogs. Biol J Linn Soc 91:347-359.

van der Meijden A, Vences M, Hoegg S, Boistel R, Channing A, and Meyer A. 2007b. Nuclear gene phylogeny of narrow-mouthed toads (Family: Microhylidae) and a discussion of competing hypotheses concerning their biogeographical origins. Mol Phylogenet Evol 44:1017-1030.

van Dongen W, de Laaf L, Zaal R, Moorman A, and Destrée O. 1981. The organization of the histone genes in the genome of *Xenopus laevis*. Nuc Acids Res 9:2297-2311.

Vandamme AM. 2009. Basic concepts of molecular evolution. Pp. 3-28. In Lemey P, Salemi M, and Vandamme AM, editors. The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, New York.

Vences M, Vieites DR, Glaw F, Brinkmann H, Kosuch J, Veith M, and Meyer A. 2003. Multiple overseas dispersal in amphibians. Proc R Soc Lond B 270:2435-2442.

Venditti C, and Pagel M. 2009. Speciation as an active force in promoting genetic evolution. Trends Ecol Evol 25:14-20.

Venkatesh B, Erdmann MV, and Brenner S. 2001. Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. Proc Natl Acad Sci USA 98:11382-11387

Verkuil YI, Piersma T, and Baker AJ. 2010. A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes). Mol Phylogenet Evol 57:411-416.

Vieites DR, Min M-S, and Wake DB. 2007. Rapid diversification and dispersal during periods of global warming by plethodontid salamanders. Proc Natl Acad Sci USA 104:19903-19907.

Vieites DR, Wollenberg KC, Andreone F, Köhler J, Glaw F, and Vences M. 2009. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. Proc Natl Acad Sci USA.

Vitt LJ, and Caldwell JP. 2009. Herpetology. Academic Press, Burlington, Massachusetts.

von Heijne G. 1986. Why mitochondria need a genome. FEBS Letters 198:1-4.

Vredenburg VT. 2004. Reversing introduced species effects: experimental removal of introduced fish leads to rapid recovery of declining frog. Proc Natl Acad Sci USA 101:7646-7650.

Wake MH. 1978. The reproductive biology of *Eleutherodactylus jasperi* (Amphibia, Anura, Leptodactylidae), with comments on the evolution of live-bearing systems. J Herpetol 12:121-133.

Wakeley J. 1994. Substitution rate varition among sites and the estimation of transition bias. Mol Biol Evol 11:436-442.

Walberg MW, and Clayton DA. 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nuc Acids Res 9:5411-5421.

Walkowiak W, and Münz H. 1985. The significance of water surface-waves in the communication of fire-bellied toads. Naturwissenschaften 72:49-51.

Wallace DC. 2005. A mitochondrial paradigm of metabolic and degenerative diseases, ageing, and cancer: a dawn for evolutionary medicine. Annu Rev Genet 39.

Wang H-C, Susko E, and Roger AJ. 2011. Fast statistical tests for detecting heterotachy in protein evolution. Mol Biol Evol 28:2305-2315.

Wang Y, and Evans SE. 2006. A new short-bodied salamander from the Upper Jurassic/ Lower Cretaceous of China. Acta Palaeontol Pol 51:127-130.

Weber R. 1974. Comparative studies on the bioacoustics of *Discoglossus pictus* Otth 1837 and *D. sardus* Tschudi 1837 (Discoglossidae, Anura). Zool Jb Physiol 18:40-84.

Webster AJ, Payne RJH, and Pagel M. 2003. Molecular phylogenies link rates of evolution and speciation. Science 301:478.

Welch JJ, and Bromham L. 2005. Molecular dating when rates vary. Trends Ecol Evol 20:320-327.

Weygoldt P. 1976. Observations on the biology and ethology of *Pipa (Hemipipa) carvalhoi* Mir. Rib. 1937. Z Tierpsychol 40:80-99.

Whitaker SL, and Knox BE. 2004. Conserved transcriptional activators of the *Xenopus* rhodopsin gene. J Biol Chem 279:49010-49018.

Wiens JJ, and Titus TA. 1991. A phylogenetic analysis of *Spea* (Anura: Pelobatidae) Herpetologica 47:21-28.

Wiens JJ. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Syst Biol 52:528-538.

Wiens JJ, Fetzner JW, Parkinson CL, and Reeder TW. 2005. Hylid frog phylogeny and sampling strategies for speciose clades. Syst Biol 54:778-807.

Wiens JJ, and Moen DS. 2008. Missing data and the accuracy of Bayesian phylogenetics. J Syst Evol 46:307-314.

Wiens JJ, and Morrill MC. 2011. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. Syst Biol 60:719-731.

Wiley EO, Johnson GD, and Dimmick WW. 2000. The interrelationships of acanthomorph fishes: a total evidence approach using molecular and morphological data. Biochem Syst Ecol 28:319-350.

Wiley EO, and Lieberman BS. 2011. Phylogenetics: the theory of phylogenetic systematics. John Wiley & Sons Inc., Hoboken, New Jersey.

Wolstenholme DR. 1992. Animal mitochondrial DNA: structure and evolution. Int Rev Cytol 141:173-216.

Wu CI, and Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc Natl Acad Sci USA 82:1741-1745.

Wu J, and Susko E. 2011. A test for heterotachy using multiple pairs of sequences. Mol Biol Evol 28:1661-1673.

Wyman SK, Jansen RK, and Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252-3255.

Xu W, Jameson D, Tang B, and Higgs P. 2006. The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. J Mol Evol 63:375-392.

Yager DD. 1992. A unique sound production mechanism in the pipid anuran *Xenopus borealis*. Zool J Linn Soc 104:351-375.

Yager DD. 1996. Sound production and acoustic communication in *Xenopus borealis*. Pp. 121-141. In Tinsley RC, and Kobel HR, editors. The Biology of *Xenopus*. Clarendom Press, Oxford.

Yamanoue Y, Miya M, Matsuura K, Sakai H, Katoh M, and Nishida M. 2009. Unique patterns of pelvic fin evolution: A case study of balistoid fishes (Pisces: Tetraodontiformes) based on whole mitochondrial genome sequences. Mol Phylogenet Evol 50:179-189.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39:306-314.

Yang Z. 1995. Evaluation of several methods for estimating phylogenetic trees when substitution rates differ over nucleotide sites. J Mol Evol 40:689-697.

Yang Z, Goldman N, and Friday AE. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation. Syst Biol 44:384-399.

Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol Evol 11:367-372.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555-556.

Yang Z, and Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Mol Biol Evol 14:717-724.

Yang Z. 1998a. Likelihood ratio test for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568-573.

Yang Z. 1998b. On the best evolutionary rate for phylogenetic analysis. Syst Biol 47:125-133.

Yang Z, and Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409-418.

Yang Z, and Rannala B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. Syst Biol 54:455-470.

Yang Z. 2006. Computational molecular evolution. Oxford University Press, New York.

Yang Z, and Rannala B. 2012. Molecular phylogenetics: principles and practice. Nat Rev Genet 13:303-314.

Yoder AD, and Zhang ZH. 2000. Estimation of primate speciation dates using local molecular clocks. Mol Biol Evol 17:1081-1090.

Yovanovich C, Jungblut L, Heer T, Pozzi A, and Paz D. 2009. Amphibian larvae and zinc sulphate: a suitable model to study the role of brain-derived neurotrophic factor (BDNF) in the neuronal turnover of the olfactory epithelium. Cell Tissue Res 336:1-9.

Zardoya R, Garrido-Pertierra A, and Bautista JM. 1995a. The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*. J Mol Evol 41:942-951.

Zardoya R, Villalta M, López-Pérez MJ, Garrido-Pertierra A, Montoya J, and Bautista JM. 1995b. Nucleotide sequence fo the sheep mitochondrial DNA D-loop and its flanking tRNA genes. Curr Genet 28:94-96.

Zardoya R, and Meyer A. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol Biol Evol 13:933-942.

Zardoya R, and Meyer A. 2000. Mitochondrial evidence on the phylogenetic position of caecilians (Amphibia: Gymnophiona). Genetics 155:765-775.

Zardoya R, and Meyer A. 2001. On the origin of and phylogenetic relationships among living amphibians. Proc Natl Acad Sci USA 98:7380-7383.

Zardoya R, Málaga-Trillo E, Veith M, and Meyer A. 2003. Complete nucleotide sequence of the mitochondrial genome of a salamander, *Mertensiella luschani*. Gene 317:17-27.

Zhang J. 1999. Performance of likelihood ratio tests of evolutionary hypotheses under inadequate substitution models. Mol Biol Evol 16:868-675.

Zhang JF, Nie LW, Wang Y, and Hu LL. 2009. The complete mitochondrial genome of the large-headed frog, *Limnonectes bannaensis* (Amphibia: Anura), and a novel gene organization in the vertebrate mtDNA. Gene 442:119-127.

Zhang P, Zhou H, Chen YQ, Liu YF, and Qu LH. 2005a. Mitogenomic perspectives on the origin and phylogeny of living amphibians. Syst Biol 54:391-400.

Zhang P, Zhou H, Liang D, Liu YF, Chen YQ, and Qu LH. 2005b. The complete mitochondrial genome of a tree frog, *Polypedates megacephalus* (Amphibia: Anura: Rhacophoridae), and a novel gene organization in living amphibians. Gene 346:133-143.

Zhang P, and Wake DB. 2009a. A mitogenomic perspective on the phylogeny and biogeography of living aecilians (Amphibia: Gymnophiona) Mol Phylogenet Evol 53:479-491

Zhang P, and Wake DB. 2009b. Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes Mol Phylogenet Evol 53:492-508

Zhong B, Deusch O, Goremykin VV, Penny D, Biggs PJ, Atherton RA, Nikiforova SV, and Lockhart PJ. 2011. Systematic error in seed plant phylogenomics. Genome Biol Evol 3:1340-1348.

Zhou Y, Zhang JY, Zheng R-Q, Yu B-G, and Yang G. 2009. Complete nucleotide sequence and gene organization of the mitochondrial genome of *Paa spinosa* (Anura: Ranoidae). Gene 447:86-96.

Zouros E, Ball AO, C. S, and Freeman KR. 1994. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. Proc Natl Acad Sci USA 91:7463-7467.

Zuckerkandl E, and Pauling L. 1962. Molecular disease, evolution, and genetic heterogeneity. Pp. 189-225. In Kasha M, and Pullman B, editors. Horizons in Biochemistry. Academic Press, New York.

Zuckerkandl E, and Pauling L. 1965. Evolutionary divergence and convergence in proteins. Pp. 97-166. In Bryson V, and Vogel HJ, editors. Evolving genes and proteins. Academic Press, New York.

APPENDICES

Appendix I

Taxon sampling, specimen vouchers, and GenBank accesion numbers.

Table A.1. Specimen vouchers and available information on collection localities for the samples used in this work. Abbreviations: ACZC; Zoological collection of Angelica Crottini, Italy; MNCN/ADN, DNA and tissue collection, Museo Nacional de Ciencias Naturales, Spain; IABH, Institute for Amphibian Biology of Hiroshima, Japan; RM, Redpath Museum, Canada; ZCMV, Zoological collection of Miguel Vences, Germany.

Species	Specimen voucher	Collection locality
Leiopelma archeyi	RM2215	Whareorino forest, west of Te Kuiti, New Zealand
Ascaphus truei	MNCN/ADN 28468	Flathead Creek, Glacier National Park, Montana, USA
Bombina orientalis	MNCN/ADN 4314	unknown locality
Discoglossus galganoi	MNCN/ADN 4315	Reliegos, Spain
Alytes dickhillenii	MNCN/ADN 28461	Spain
Rhinophrynus dorsalis	MNCN/ADN 28469	Tenexpa, Guerrero, Mexico
Pipa carvalhoi	MNCN/ADN 28466	unknown locality
Xenopus laevis	MNCN/ADN 28464	Jonkershoek, South Africa
Hymenochirus boettgeri	MNCN/ADN 28465	unknown locality
Pseudhymenochirus merlini	MNCN/ADN 28467	bred in captivity, parents from *ca.* 130 Km East of the capital Bissau, Guinea Bissau
Pelobates fuscus fuscus	ACZC0053	Turin, Italy
Pelodytes punctatus	MNCN/ADN 8000	Portalegre, Portugal
Heleophryne regis	MNCN/ADN 28481	unknown locality
Lechriodus melanopyga	MNCN/ADN 8001	Papua New Guinea
Calyptocephalella gayi	MNCN/ADN 8002	Chile
Telmatobius bolivianus	MNCN/ADN 563	Sud Yungas, La Paz, Bolivia
Sooglossus thomasseti	MNCN/ADN 28482	unknown locality
Sooglossus sechellensis	MNCN/ADN 28483	unknown locality
Duttaphrynus melanostictus	ZCMV11016	unknown locality
Hyla chinensis	ZCMV11019	unknown locality
Microhyla sp.	MNCN/ADN 28462	unknown locality
Kaloula pulchra	ZCMV11017	unknown locality
Fejervarya limnocharis	MNCN/ADN 28470	Sri Lanka
Mantella madagascariensis	IABH6960	Sri Lanka
Polypedates cruciger	MNCN/ADN 28463	Sri Lanka
Rhacophorus dennnysi	ZCMV11011	Sri Lanka

Table A.2. Taxon sampling and GenBank accession numbers of sequences used in the different analyses, including the phylogenetic analyses in all three studies, and divergence time estimation. Newly generated data are shown in bold. For some species, only mitochondrial genome data is available, but a nuclear matrix with information from nuclear loci was generated by sequencing partial nuclear loci and using available data from GenBank. In few cases, chimerical sequences were generated by using data from closely related species (mostly within the same genus). In that case, the species is shown below the corresponding GenBank accesion number.

	SPECIES	Mitochondrial genome	*bdnf*	*cxcr4*	*h3a*	*pomc*	*rag1*	*rag2*	*rho*	*slc8a1*	*slc8a3*
Mammalia	*Homo sapiens* (human)	D38112	NM_001143816	NM_001008540	NM005324	NM_000939.2	NM_000448	NG_007573	U49742	X91213	X93017
Aves	*Gallus gallus* (bird)	X52392	NM_001031616	NM204617	Y00392	GU269642	M58530	AY443150	D00702	NM_001079473	XM421178
Sauropsida	*Iguana iguana* (lizard)	AJ278511	FJ433956 *Cyclura nubila*	EF110995 *Timon lepidus*	DQ284249	AB128826 *Eublepharis macularius*	EU402826.1 *Anolis carolinensis*	DQ119641 Cyclura nubila	AY902453 *Anolis carolinensis*	GU456076 *Anolis carolinensis*	EF110997 *Timon lepidus*
Cryptobranchidae	*Andrias davidianus*	AJ492192	EU275889	AY948801	DQ284358	EU275843	AY583346 *A. japonicus*	AF369085 *Pachytriton* sp.	—	AY948847	AY948911
Hynobiidae	*Batrachuperus pinchonii*	DQ333815	EU275864 *Hynobius sonani*	EF017998	DQ284330	EU275818 *H. sonani*	EF018054	—	DQ347406 *B. formosanus*	EF018023	EF107362
	Ranodon sibiricus	NC_004021	—	—	—	—	—	—	—	—	—
Sirenidae	*Siren intermedia*	GQ368661	—	EF107473	DQ284216 *S. lacertina*	—	AY650140	—	—	EF107249	EF107407
Rhinatrematidae	*Rhinatrema bivittatum*	AY456252	—	EF107478	DQ284370	—	AY456257	—	DQ284002	EF107255	EF107417
Ichthyophiidae	*Ichithyophis glutinosus*	AY456251	—	AY948794	DQ284137 *I.* cf. *peninsualirs*	—	AY456256	—	—	AY948839	AY948901
Typhlonectidae	*Typhlonectes natans*	AF154051	—	—	DQ284136	AF369043	AY456260	AF369088	—	EF107229 *Typhlonectes* sp.	EF107365 *Typhlonectes* sp.
Leiopelmatidae	*Ascaphus truei*	AJ871087	EU275896	AY523695	DQ284162	EU275850	AY323754	**HM998977**	AY323730	AY523731	AY948893
	Leiopelma archeyi	**HM142901**	**HM998927**	AY523700	**HM998942**	**HM998959**	**HM998973**	**HM998978**	DQ283895	**HM998951**	EF107408
Bombinatoridae	*Bombina orientalis*	AY585338	**HM998928**	AY364177	**HM998943**	AY692246	AY583335	AY323783	**HM998984**	AY523715	AY948867
	Bombina bombina	—	—	—	—	—	—	—	—	—	—
	Bombina maxima	—	—	—	—	—	—	—	—	—	—
	Bombina variegata	—	—	—		—	—	—	—	—	—
Alytidae	*Alytes obstetricans*	AY585337	EF407511	AY364170	**HM998945** *A. dickhilleni*	**HM998961** *A. dickhilleni*	AY583334	AY323781 *A. muletensis*	DQ283825	Y523703	EF107345
	Discoglossus galganoi	AY585339	**HM998929**	AY364172 *D. pictus*	**HM998944**	**HM998960**	AY583338	AY323785 *D. sardus*	DQ283915	AY523708 *D. pictus*	AY948858 *D. pictus*
Rhinoprhynidae	*Rhinophrynus dorsalis*	**HM991334**	**HM998933**	AY523699	**HM998946**	**HM998962**	AY874302	**HM998979**	DQ347405	AY523722	AY948894
Pipidae	*Pipa pipa*	GQ244477	—	AY364174	—	—	AY874303	—	DQ283781	AY523711	EF107351
	Pipa carvalhoi	**HM991332**	**HM998935**	—	DQ284277	**HM998963**	**HM998974**	**HM998980**	DQ283922	**HQ260711**	—
	Silurana tropicalis	AY789013	EF433430	AY523697	CR855729	BC088054	AY874306	EF535957	NM_001097334	AY523721	AY948891
	Xenopus laevis	**HM991335**	**HM998930**	AY523691 *Xenopus* sp.	J00984	X05941	L19324	L19325	S62229	X90839	EF107370 *X. wittei*
	Hymenochirus boettgeri	**HM991331**	**HM998932**	AY523685	**HM998947**	**HM998964**	AY583340	**HM998981**	AY323735	AY523702	EF107344
	Pseudhymenochirus merlini	**HM991333**	**HM998934**	—	**HM998948**	**HM998965**	**HM998975**	**HM998982**	**HM998985**	**HM998953**	—

Pelobatidae	*Pelobates cultripes*	AJ871086	**HM998931** *P. fuscus*	AY364171	DQ284159 *P. fuscus*	**HM998966** *P. fuscus*	AY323758	**HM998983** *P. fuscus*	AY323736	AY523707	AY948857
Pelodytidae	*Pelodytes punctatus*	**JF703231**	**JF703235**	—	DQ284157	**JF703242**	AY583343	**JF703247**	DQ283824	AY523709	—
Heleophrynidae	*Heleophryne regis*	**JF703229**	**JF703237**	AY364191 *H. purcelli*	DQ284161	**JF703243**	AY323764	AY323786	AY323739	AY948833 *H. purcelli*	AY948892 *H. purcelli*
Calyptocephalellidae	*Calyptocephalella gayi*	**JF703228**	**JF703236**	EF107495	DQ284415	AY819090	AY583337	DQ872909	DQ284036	EF107275	EF107440
Limnodynastidae	*Lechriodus melanopyga*	**JF703230**	**JF703238**	-	DQ284299 *L. fletcheri*	**JF703244**	AY583341	DQ872908	DQ283942 *L. fletcheri*	**JF703249**	—
Hylidae	*Hyla chinensis*	AY458593	**HM998936**	AY523687 *H. meridionalis*	**HM998949**	DQ055794	HM998976	HQ260710	AY844615 *H. japonica*	**HM998954**	AY948860
	Hyla japonica	NC_010232	—	—	—	—	—	—	—	—	—
Bufonidae	*Duttaphrynus melanostictus*	AY458592	**HM998937**	AY364167	DQ284324	DQ158317	EU712821	AY323784 Amietophrynus regularis	DQ283967	AY948805	AY948851
	Bufo japonicus	NC_009886	—	—	—	—	—	—	—	—	—
	Bufo gargarizans	NC_008410	—	—	—	—	—	—	—	—	—
Ceratophrydae	*Telmatobius bolivianus*	**JF703234**	**JF703241**	EF107464 Telmatobius sp.	DQ284068 *T. marmoratus*	AY819097 *T. truebae*	AY583344	**JF703248**	DQ283770 *T. verrucosus*	EF107239 *Telmatobius* sp.	EF107389 *Telmatobius* sp.
Sooglossidae	*Sooglossus thomasseti*	**JF703233**	**JF703239**	AY364187	DQ284425	**JF703245**	AY323778	AY323798	AY323744	**JF703250**	AY948884
	Sooglossus sechellensis	**JF703232**	**JF703240**	-	DQ284423	**JF703246**	DQ872921	DQ872910	DQ284040	**JF703251**	—
Dicroglossidae	*Fejervarya limnocharis*	AY158705	**HM998938**	-	DQ284356	**HM998969**	AY571649 Fejervarya sp.	DQ019526 Fejervarya sp.	DQ458271	**HM998955**	—
	Fejervarya cancrivora	NC_012647	—	—	—	—	—	—	—	—	—
	Quasipaa spinosa	NC_013270	—	—	—	—	—	—	—	—	—
	Limnonectes fujianensis	NC_007440	—	—	—	—	—	—	—	—	—
	Limnonectes bannaensis	NC_012837	—	—	—	—	—	—	—	—	—
Mantellidae	*Mantella madagascariensis*	AB212225	**HM998940**	-	DQ284061 *M. aurantiaca*	**HM998971**	DQ019500	DQ019532	AY263284	**HQ260711**	—
Rhacophoridae	*Polypedates megacephalus*	AY458598	**HM998939** *P. cruciger*	-	DQ284079 *P. leucomystax*	**HM998970** *P. cruciger*	HQ260712 *P. cruciger*	AY323802 *P. maculatus*	EU924545	**HM998956**	—
	Rhacophorus schlegelii	AB202078	**HM998941** *R. dennysi*	AY948769 *R. malabaricus*	**HM998950** *R. dennysi*	**HM998972**	DQ019512 *R. dennysi*	DQ019547 *R. dennysi*	EU215575 *R. dennysi*	**HM998958**	AY948848 *R. malabaricus*
	Buergeria buergeri	NC_008975	—	—	—	—	—	—	—	—	—
Microhylidae	*Kaloula pulchra*	AY458595	EF396015	EF017974	DQ284379	**HM998968**	AY323772	AY323790	DQ284011	EF018030	AY948853
	Microhyla ornata	DQ512876	EF396021 *M. pulchra*	AY364168	DQ284400 *Microhyla* sp.	**HM998967** *Microhyla* sp.	EF396093 *Microhyla* sp.	EF396134 *M. pulchra*	AY364383	AY948806	AY948852
	Microhyla heymonsi	NC_006406	—	—	—	—	—	—	—	—	—
	Micorhyla okinavensis	NC_010233	—	—	—	—	—	—	—	—	—
Ranidae	*Odorrana tormota*	NC_009423	—	—	—	—	—	—	—	—	—
	Pelophylax nigromaculatus	AB043889	—	—	—	—	—	—	—	—	—
	Pelophylax plancyi	NC_009264	—	—	—	—	—	—	—	—	—

Appendix II

Results from the relative-rate tests for amino acid data.

Table. A.2. Results from relative rate tests based on amino acid data of single genes and combined mt and nuclear datasets. Mean weighted substitution rates (K) for (1) non-neobatrachians, (2) all Neobatrachia, (3) basal neobatrachians (*Heleophryne, Calyptocephalella*, *Lechriodus*, and *Sooglossus*), and (4) derived neobatrachians (Ranoides and Nobleobatrachia). Probability values (p) of relative-rate tests are shown for each comparison, with corresponding groups in parentheses. Statistically significant results ($p < 0.05$; or $p < 0.5/3 = 0.0167$ after Bonferroni correction) are in bold italics and marked with an asterisk.

Gene	K1	K2	K3	K4	p (1 vs. 2) p<0.05	p (1 vs. 3) p<0.00167	p (1 vs. 4) p<0.00167	p (3 vs. 4) p<0.00167
atp6	0.330015	0.367739	0.374928	0.365536	0.16137	0.112428	0.185803	0.685741
atp8	0.714764	0.826892	0.822232	0.823635	0.229052	0.263528	0.248108	0.987666
cob	0.257653	0.285157	0.293829	0.240293	0.436551	0.352768	0.611322	0.218300
cox1	0.052609	0.095683	0.097475	0.099125	***3.35·10^{-6}****	***2.51·10^{-6}****	***2.05·10^{-6}****	0.824218
cox2	0.133924	0.262373	0.265594	0.247925	***9.60·10^{-7}****	***1.59·10^{-6}****	***8.95·10^{-6}****	0.395054
cox3	0.143303	0.183249	0.185461	0.179850	***0.014532***	***0.013051****	0.0312632	0.690831
nad1	0.274486	0.312906	0.313724	0.310114	***0.044270***	0.0460851	0.0465841	0.826222
nad2	0.505490	0.618186	0.614789	0.578833	***0.000158***	***0.000343****	***0.007778***	0.189552
nad3	0.359680	0.349876	0.349364	0.382862	0.736314	0.735309	0.441277	0.243505
nad4	0.369812	0.595501	0.596860	0.631774	***1.00·10^{-7}****	***1.00·10^{-7}****	***1.00·10^{-7}****	0.144633
nad4L	0.442918	0.587862	0.586601	0.591903	***0.019638****	0.0271234	***0.013444***	0.915354
nad5	0.353876	0.505448	0.488711	0.564544	***0.0000001****	***1.00E-07***	***1.00·10^{-7}****	***1.14·10^{-5}****
nad6	0.454670	0.495709	0.503102	0.519844	0.316416	0.254464	0.104917	0.556079
all mt genes	0.254548	0.332774	0.334849	0.333448	***0.0000001****	***1.00·10^{-7}****	***1.00·10^{-7}****	0.825768
bdnf	0.063766	0.081077	0.080242	0.084697	0.066639	0.0686618	0.064521	0.548784
cxcr4	0.199414	0.222181	0.222604	0.220910	0.189471	0.197844	0.224584	0.910408
h3a	0.000307	0.000101	0.000000	0,000403	0.315292	0.0836205	0.827200	0.317486
pomc	0.382952	0.420487	0.417945	0.434916	0.390285	0.431069	0.274356	0.591228
rag1	0.135477	0,136645	0.137594	0.134050	0.942025	0.897386	0.932186	0.817249
rag2	0.444987	0.510587	0.516113	0.520481	***0.019122****	***0.015075****	***0.010908****	0.844919
rho	0.128185	0.133436	0.128823	0.169018	0.635837	0.961838	0.057867	0.096293
slc8a1	0.074735	0.085632	0.084384	0.082348	0.084604	0.149769	0,213418	0.770769
slc8a3	0.088013	0.091065	0.091514	0.088908	0.671885	0.63757	0.904896	0.663627
all nuclear genes	0.092941	0.100978	0.100373	0.711942	0.218256	0.273498	0.451904	0.451904

Appendix III

Results from the branch models used to estimate selection on DNA sequences.

Table A.3. Selection coefficients (ω) and differences in the values of the Akaike information criterion (ΔAIC) for all 10 branch models tested, including (i) null model (background), and alternative models that assumed a second independent ω in (ii) the stem branch of Neobatrachia, (iii) the whole Neobatrachia, (iv), Ranoides, (v) Nobleobatrachia, (vi) Pipoidea, (vii) the stem branch of Pelobatoidea, (viii) the whole Pelobatoidea, (ix) Discoglossoidea, and (x) Amphicoela. Significant differences between alternative versus the null model (LRT; $p < 0.05$) are shown in bold, both for in ω and ΔAIC values.

Selection coefficients (ω)

Model	(i) null	(ii) Neob.-stem	(iii) Neob.-all	(iv) Ranoides	(v) Nobleob.	(vi) Pipoidea	(vii) Pelob.-stem	(viii) Pelob.-all	(ix) Discogl.	(x) Amphic.
atp6	0.031	0.036	0.031	0.042	1.123	0.028	0.038	0.033	0.033	0.030
atp8	0.148	1.508	0.167	0.354	0.136	**0.101**	0.999	0.128	0.100	0.225
cob	0.034	**0.069**	0.034	0.043	0.044	0.030	0.041	0.035	**0.026**	0.036
cox1	0.011	**0.037**	**0.014**	**0.039**	0.024	**0.006**	0.014	0.009	**0.005**	0.012
cox2	0.022	0.026	0.023	0.034	0.024	0.021	0.026	0.027	**0.013**	0.023
cox3	0.027	**0.088**	**0.030**	0.058	0.039	**0.021**	0.041	0.029	0.021	0.028
nad1	0.030	**0.051**	**0.032**	0.043	0.034	**0.025**	0.033	0.028	0.030	0.032
nad2	0.028	0.033	0.028	0.030	0.028	0.028	0.029	0.028	0.028	0.028
nad3	0.065	0.144	0.066	0.932	0.105	0.065	0.055	0.052	0.065	0.076
nad4	0.036	0.039	0.037	0.037	0.041	0.034	0.039	0.037	0.036	0.037
nad4L	0.045	0.048	0.043	0.046	0.043	0.049	0.059	0.043	0.041	0.058
nad5	0.035	0.038	0.036	0.038	0.037	0.034	0.031	0.033	0.034	0.037
nad6	0.029	0.031	0.031	0.034	0.033	0.027	0.027	0.026	0.026	0.030
all mt genes	0.047	**0.083**	**0.049**	**0.071**	**0.064**	**0.042**	0.056	0.046	**0.042**	0.049
bdnf	0.046	0.012	0.050	0.112	0.054	0.045	0.056	0.031	0.035	0.088
cxcr4	0.065	0.067	0.073	0.075	0.031	**0.040**	**0.003**	0.043	**0.104**	0.091
h3a	0.005	0.000	0.003	0.000	0.000	0.009	0.000	0.006	0.003	0.000
pomc	0.082	0.236	0.070	**0.026**	0.057	0.087	0.028	0.075	0.109	**0.149**
rag1	0.057	0.037	**0.065**	**0.017**	0.044	**0.068**	**0.021**	**0.037**	0.048	0.083
rag2	0.160	0.249	0.152	0.046	0.000	0.171	0.124	0.137	0.173	0.308
rho	0.090	0.024	**0.114**	0.050	0.027	0.074	0.051	0.055	0.060	0.183
slc8a1	0.040	**0.000**	**0.047**	0.015	0.032	0.032	0.046	0.044	0.048	**0.111**
slc8a3	0.026	0.014	0.027	0.010	0.020	0.024	0.035	0.019	0.023	**0.075**
all nuclear genes	0.069	0.062	**0.077**	**0.033**	**0.042**	0.068	**0.045**	**0.054**	0.070	**0.115**

ΔAIC values

atp6	0	2	2	2	2	1	2	2	2	2
atp8	5	4	4	6	7	0	7	7	4	5
cob	7	0	9	9	9	7	9	9	3	9
cox1	42	27	0	36	43	11	44	43	27	44
cox2	6	7	8	7	8	8	8	6	0	8
cox3	5	0	2	7	7	1	5	7	5	7
nad1	3	1	1	4	5	0	5	5	5	5
nad2	0	2	2	2	2	2	2	2	2	2
nad3	0	0	2	2	1	2	2	0	2	1
nad4	0	2	2	2	2	1	2	2	2	2
nad4L	0	2	1	2	2	1	2	2	2	1
nad5	0	1	1	2	2	2	2	1	2	2
nad6	0	2	0	2	2	1	2	2	2	2
all mt genes	78	0	56	61	68	48	77	80	66	79
bdnf	1	0	2	2	3	3	3	2	2	1
cxcr4	10	12	10	12	9	0	4	9	4	11
h3a	1	2	1	2	1	0	2	3	2	2
pomc	7	5	5	0	8	8	8	8	7	5
rag1	8	8	3	4	9	6	4	0	8	5
rag2	7	7	7	7	9	8	8	7	8	0
rho	3	3	0	3	3	4	4	2	3	1
slc8a1	15	10	12	13	16	14	17	16	15	0
slc8a3	15	16	17	15	16	16	16	14	16	0
all nuclear genes	32	33	14	18	24	34	28	19	34	0

Appendix IV

Results from the analysis of molecular synapomorphies in Neobatrachia

Table. A.4. Results from the analyses performed with MrFunction. The number of synapomorphic sites (No. sites) is shown both for Neobatrachia and Pelobatoidea in all mitochondrial and nuclear genes. The difference in the number of synapomorphic changes between both groups is tested using a binomial test, and significant differences ($p < 0.05$) are shown in bold italics.

	No. sites	Neobatrachia	Pelobatoidea	Probability (p) Binomial test
Mitochondrial genes				
atp6	13	7	6	1
atp8	1	1	0	1
cob	12	12	0	***0.000***
cox1	12	6	6	1
cox2	8	6	2	0.289
cox3	10	2	8	0.109
nad1	9	2	7	0.180
nad2	4	3	1	0.625
nad3	3	1	2	1
nad4	12	7	5	0.774
nad4L	2	0	2	0.5
nad5	57	53	4	$\boldsymbol{5.91 \cdot 10^{-12}}$
nad6	8	2	6	0.289
Nuclear genes				
bdnf	1	0	1	1
cxcr4	10	2	8	0.109
h3a	0	0	0	—
pomc	5	4	1	0.375
rag1	9	6	3	0.508
rag2	14	12	2	***0.013***
rho	1	0	1	1
slc8a1	1	0	1	1
slc8a3	5	0	5	0.063
Total	**197**	**126**	**71**	


mncn
museonacionaldecienciasnaturales


UAM
UNIVERSIDAD AUTONOMA
DE MADRID